Investigation of Novel Structures and Functions of Readers and Writers of the Histone Code in *Arabidopsis thaliana*

Thesis submitted for the degree of *Philosophiae Doctor*

by

Mohummad Aminur Rahman

Department of Molecular Biosciences
Faculty of Mathematics and Natural Sciences
University of Oslo, 2012

Series of dissertations submitted to the
Faculty of Mathematics and Natural Sciences, University of Oslo
No. 1185

ISSN 1501-7710

Cover: Inger Sandved Anfinsen.
Printed in Norway: AIT Oslo AS.

Produced in co-operation with Unipub.
The thesis is produced by Unipub merely in connection with the
thesis defence. Kindly direct all inquiries regarding the thesis to the copyright
holder or the unit which grants the doctorate.

Table of Contents

Acknowledgements

The work presented in this thesis has been carried out at the Department of Molecular Biosciences, University of Oslo and financially supported by the faculty of Mathematics and Natural Sciences, University of Oslo.

First I would like to thank my supervisor professor Reidunn B. Aalen for giving me the opportunity to do my Ph.D. in her group, for all help and support during this project, especially for the invaluable help in the writing process, and for always having time for discussions. I would also like to thank my co-supervisor Dr. Tage Thorstensen for all the help during this period. From the deep of my heart, I am remembering Dr. William Davis, my second co-supervisor, who is not with us anymore and hoping that his soul is in peace.

I would also express my gratitude to Dr. Per E. Kristiansen for helping me in the NMR project. Without your help it would be tough for me to learn NMR and solve the structure of WIYLD domain.

Especially I would like to thank Silje V. Veiseth for all the help and discussion in the laboratory. I would like to thank Robert Kumpf for the valuable discussion during the writing process. Thanks to the people of Ute Krengel's group in the Department of Chemistry, UiO for helping me during the crystallography part of the project. Solveig H. Engebretsen, Roy Falleth and Maren H. Backstrøm, thank you for all your technical assistance on this project. Thanks to all the people at the lab for making a nice work place together.

I would like to thank all my family and friends for all the support, time and for always being there for me. Thanks to the members of 'Oslo Tigers Cricket Club' to make my time more enjoyable in Oslo.

Finally and most importantly, I wish to thank my beautiful wife Firoza Begum for always being supportive and patient with me during these years. Without your support it would be impossible for me to finish this. Thank you Adyan, my lovely son, to make my life complete, and I love you with all my heart!

Abbreviations

ASHH	ASH1 HOMOLOG
ATP	Adenosine 5'-triphosphate
Col	Columbia
ChIP	Chromatin immunoprecipitation
CSP	Chemical shift perturbation
DNA	Deoxy ribonucleic acid
ELISA	Enzyme-linked immunosorbent assay
GFP	Green fluorescent protein
H2Bub1	H2B monoubiquitinated on lysine 143
H3K	Histone H3 lysine
HKMTases	Histone lysine methyltransferases
NMR	Nuclear magnetic resonance
OE	Overexpression
PTM	Post-translational modification
UBD	Ubiquitin binding domain
SUVR	SU(VAR) 3-9 RELATED
UBQ1	UBIQUITIN EXTENTION PROTEIN 1
Y2H	Yeast two-hybrid

List of the papers

Paper I - PLoS Genetics 7(3): e1001325, 2011

The SUVR4 Histone Lysine Methyltransferase Binds Ubiquitin and Converts H3K9me1 to H3K9me3 on Transposon Chromatin in *Arabidopsis*

*Silje V. Veiseth, **Mohummad A. Rahman**, Kyoko L. Yap, Andreas Fischer, Wolfgang Egge-Jacobsen, Gunter Reuter, Ming-Ming Zhou, Reidunn B. Aalen, and Tage Thorstensen.*

Paper II - Manuscript

NMR Analysis of the WIYLD Domain of the *Arabidopsis* Histone Methyltransferase SUVR4 Reveals a Novel Ubiquitin-Binding Four-Helix Bundle Structure

***Mohummad A. Rahman**, Per E. Kristiansen, Silje V. Veiseth, Jan Terje Andersen, Kyoko L. Yap, Inger Sandlie, Tage Thorstensen and Reidunn B. Aalen.*

Paper III - EMBO Journal 30(10): 1939-1952, 2011

The CW domain, a new histone recognition module in chromatin proteins

*Verena Hoppmann, Tage Thorstensen, Per Eugen Kristiansen, Silje Veie Veiseth, **Mohummad Aminur Rahman**, Kenneth Finne, Reidunn B Aalen, and Rein Aasland.*

1. Introduction

A major challenge for eukaryotic organisms is to fit their large genome into a small nucleus and to organize the DNA as to facilitate transcription, replication and repair within this limited volume. This challenge is overcome by assembly of genomic DNA into chromatin, which is a complex of DNA, histones and non-histone proteins. Chromatin is a dynamic structure that can exist as condensed chromatin fibers constituting heterochromatin, and decondensed euchromatin. Condensed, constitutive heterochromatin is in general transcriptionally repressive, has a low gene density, but is rich in repetitive elements, while the more open euchromatin is gene-rich and may allow transcription. Heterochromatin, euchromatic transcriptionally active genes and silent genes differ in their content of covalent post-translational modifications (PTMs) on the histones. Combinations of PTMs may constitute a histone code specifying chromatin states (Strahl and Allis, 2000; Jenuwein and Allis, 2001; Lee et al., 2010). The PTMs are thought to create binding sites for chromatin-associated sensor and effector proteins facilitating or restricting transcription. While histone acetylation is generally associated with permissive chromatin, methylation of different amino acid residues correlate with permissive, active or repressed chromatin. Methylation of lysine residues on the histone tails is conferred by histone lysine methyltransferases (HKMTases) which have a so called SET domain, responsible for the enzymatic function (Jenuwein et al., 1998; Rea et al., 2000). In addition to the SET domain, several conserved additional domains (co-domains) are present in SET domain proteins, many with unknown function. Co-domains may contribute to the recruitment of histone modifiers to relevant sites in chromatin and/or may modulate the activity of the methyltransferases. For example, the chromo domain found in the members of the SU(VAR)3-9 family in *Drosophila*, yeast, insect and mammalian HMTase proteins, is recognizing methylated lysine and leading to epigenetic repression (Schotta et al., 2002; Komander et al., 2009). In this thesis, the focus has been to characterize co-domains of selected SET domain proteins in the model plant *Arabidopsis thaliana* by functional studies investigated by different epigenetic methods and 3D structure analysis by NMR.

1.1. Histones and post-translational modifications

Chromatin is the combination of DNA and proteins found in the nucleus of eukaryotic cells. Its main functions are to package DNA as to fit into the small volume of the nucleus, to structure the DNA to allow mitosis and meiosis, to protect DNA against damage and to control gene expression and DNA replication. The fundamental unit of chromatin is the nucleosome and the building blocks of nucleosomes are four core histones (H3, H4, H2A, H2B) (Figure 1 and 2). Two of each of the core histones assemble to form one octameric nucleosome core particle, with 147 base pairs of DNA wrapped around in a 1.7 left-handed super helical turn (Luger et al., 1997; Quina et al., 2006; Robinson et al., 2006).

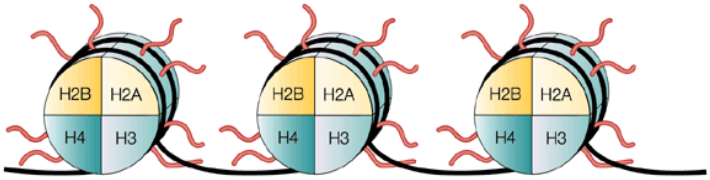

Figure 1. The core proteins of nucleosomes are designated histone H2A, histone H2B, histone H3 and histone H4 (figure from Marks et al. (2001)). Each histone is present in two copies, and the DNA (black) wraps around an octamer of histones - the core nucleosome.

The nucleosomes appear as a 'beads on a string' structure connected by 10-60 bp linker DNA, which constitutes a chromatin fiber that is 10 nm in diameter (Horn and Peterson, 2002; Robinson et al., 2006). This 10-nm fibre folds into a chromatin structure of higher order, and coil into shorter and thicker fiber of 30 nm in diameter (Figure 2). The higher order structure is stabilized by linker histone H1 (Quina et al., 2006; Robinson et al., 2006).

Two major forms of chromatin with different degree of condensation of the chromatin fiber are found in interphase nuclei. The lightly packed form of chromatin is euchromatin, which is rich in gene content and facilitates active transcription. In contrast, heterochromatin, found around centromeres, transposable elements and telomeres, is a tightly packed form, repressing transcriptional activity. Multiple epigenetic mechanisms, involving the post-translational modification of the N-terminal tails of core histones on the specific residues (Figure 2 and 3), DNA methylation, ATP-dependent chromatin remodeling, placement of histone variants, and noncoding RNA, are regulating the structure and function of chromatin (Kouzarides, 2007; Latham and Dent, 2007; Liu et al., 2010).

Each of the histone unit has a basic N-terminal tail protruding from the nucleosome that is involved in inter-nucleosomal histone-histone interactions, and contains sites for PTMs on different amino acids, such as acetylation, methylation, phosphorylation and ubiquitination on lysine (K), arginine (R) or serine (S) residues (Figure 2 and 3). Different combinations of these covalent modifications seem to be crucial for turning specific genes on or off, i.e. activating or repressing transcription (Peterson and Laniel, 2004; Latham and Dent, 2007). The PTMs might provide specificity for effector proteins that bind the modification marks and interpret this into functional outcomes (Jenuwein and Allis, 2001). It has also shown that transcription, mitosis and chromosome stability can be affected by these modifications (Strahl and Allis, 2000).

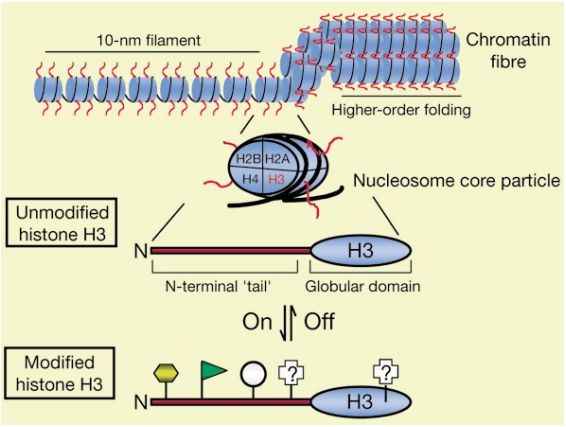

Figure 2. General chromatin organization (figure from Strahl and Allis (2000)). Like other histone 'tails', the N terminus of H3 (red) represents a highly conserved domain that is likely to be exposed or extend outwards from the chromatin fiber. A number of distinct post-translational modifications are known to occur at the N-terminus of H3 including acetylation (green flag), phosphorylation (grey circle) and methylation (yellow hexagon). Other modifications are known and may also occur in the globular domain.

Acetylation of histone tail lysine residues on histone H3, H4, H2A and H2B, correlating with active chromatin and facilitating transcription and gene expression (Table 1) (Xu et al., 2005; Berger, 2007; Kouzarides, 2007), is catalysed by histone acetyltransferase and reversed by histone deacetylases (HDACs) (Khorasanizadeh, 2004). Acetylation is believed to promote DNA accessibility for enzymes involved in transcription (Khorasanizadeh, 2004). Recent studies suggest that nucleosomes with lysine acetylated residues usually rearrange during plant development, and acetylation modification is related to root elongation, flowering and cold tolerance (reviewed in Chen et al. (2010)).

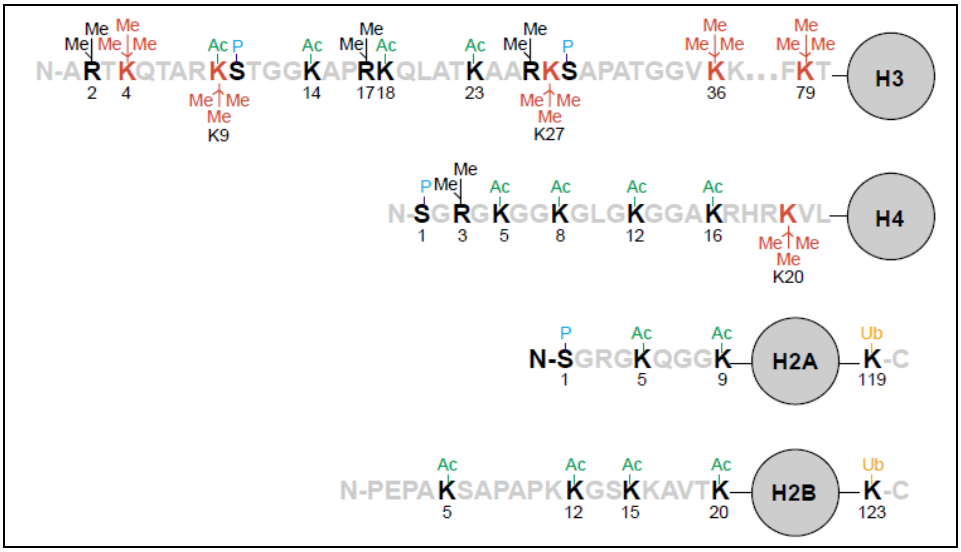

Figure 3. Schematic representation of the post-translational modifications on core histones (H2A, H2B, H3 and H4) (figure from Sims et al. (2003)). Modifications – methylation (Me), acetylation (Ac), phosphorylation (P) and ubiquitination (Ub) – shown above each amino acid – lysine (K), arginine (R) and serine (S) – correlate with activation, whereas the lysine methylation shown below correlates with repression. The highest degree of methylation possible for each residue is displayed.

Methylation of histone tail residues has a role in both activation and repression of transcription. These particular modifications are catalysed by histone methyltransferases (HMTases) and histone H3 and H4 tails have long been known as substrates (Figure 3). In general, methylation of histone H3K9 and H3K27 is required for transcriptional repression, methylation of histone H3K4 and H3K36 is required for transcriptional activation (Table 1) (Zhou, 2009). It should be noted that methylation of H3K79 and H4K20 is found in non-plant systems. In yeast, H3K79 methylation is associated with telomere silencing, meiotic checkpoint control and DNA damage response (Jones et al., 2008). In *Drosophila* methylation of H4K20 is linked with transcriptional repression (Karachentsev et al., 2005).

Serine phosphorylation on the histone tails is also associated with transcriptional activation and DNA repair (Smith and Walker, 1996). Phosphorylation occurs mainly during cell mitosis and it has been reported that alteration of H2B serine phosphorylation is connected to apoptosis (Zhang et al., 2007; Cerutti and Casas-Mollano, 2009).

Ubiquitination take place on histone H2A and H2B and this is a process of attaching the ubiquitously expressed ubiquitin to lysine residues as a post-translational modification. Poly-

ubiquitination of H2A has been reported, but monoubiquitinated K119 in H2A appears to be the major form (Nickel and Davie, 1989; Winget and Mayor, 2010). In contrast to H2Aub, H2Bub is conserved from yeast (K123) to human (K120) and plants (K143) (West and Bonner, 1980; Zhang et al., 2007). Although ubiquitination is generally associated with proteasomal degradation of proteins, it is also involved in regulation of protein activity and transcription (Ikeda and Dikic, 2008; Xu et al., 2009; Hammond-Martel et al., 2011). Several mechanisms have been provided in different studies for the regulation of transcription by H2A or H2B ubiquitination (Hammond-Martel et al., 2011). However, ubiquitination on histone H2A is required for transcription repression in human and yeast, while ubiquitination of H2B is required for transcription activation by ensuring cross-talk between histone modifications (Zhang et al., 2007; Hammond-Martel et al., 2011). It has been shown that monoubiquitination on H2A in *Arabidopsis* plays a key role in maintaining cell identity (Winget and Mayor, 2010).

Modification	Histone	Residues	Function
Acetylation	H2A	K5	Transcriptional activation
		K144	
	H2B	K6, K11, K27, K32	
	H3	K9, K14, K18, K23	Transcriptional activation
	H4	K5	Transcriptional repression
		K8, K12, K16	Transcriptional activation
		K20	
Methylation	H3	K4	Transcriptional activation
		K9, K27	Transcriptional repression
		K36	Transcriptional activation / repression
Phosphorylation	H2A	S129	Cell cycle progression (Mitosis)
		S141, S145	
		S138	DSB repair
	H2B	S15	Apoptosis
Ubiquitination	H2A	K119	Maintaining cell identity
	H2B	K143	Transcriptional activation

Table 1. Types of covalent histone post-translational modifications and their function in *Arabidopsis* (Zhang et al., 2007).

Histone can also be post-translationally modified by SUMOylation, where large ubiquitin-like moieties added to lysines on H2A, H2B and H4 seem to correlate with both euchromatic transcriptional repression and heterochromatic gene silencing (Shiio and Eisenman, 2003; Nathan et al., 2006).

1.2. SET domain proteins are histone lysine methyltransferases

Two different classes of enzymes are responsible for methylation of arginine (R) and lysine (K) residues, histone arginine methyltransferases (HRMTases) and histone lysine methyltransferases (HKMTases), respectively. The HRMTases are responsible for catalyzing the transfer of methyl groups from S-adenosyl-L-methionine over to the nitrogen atoms on the arginine residue side chains (Zhang and Reinberg, 2001). Arginine can be mono- or dimethylated, and depending on the dimethylation being symmetrical or asymmetrical, the HRMTases are divided in two different groups (Zhang and Reinberg, 2001). HKMTases have enormous specificity to histone tail lysines and they usually modify one single lysine on a single histone and their output can be either activation or repression of transcription (Bannister and Kouzarides, 2005) (Figure 3). The HKMTases contain a SET domain, which is a 130-160 amino acids evolutionary conserved domain and was first recognized in three *Drosophila melanogaster* proteins **S**UPPRESSOR OF VARIEGATION 3-9 [SU(VAR)3-9], **E**NHANCER OF ZESTE [E(Z)], AND **T**RITHORAX (TRX) (Figure 4) (Jenuwein et al., 1998). This domain has been found also in a range of proteins in different organisms, i.e. in nematodes, fungi, insects, mammals and plants. SET domain dependent methylation of histone H3K9 by SUV39H1 and Clr4 was first discovered in 2000 (Rea et al., 2000). Since then a number of SET domain proteins have been shown to process HKMTase activity. The HKMTase activity differs both regarding which lysine residue they methylate (substrate specificity) and the number of methyl groups they add to each lysine (product specificity). The lysine residues can be mono-, di- or trimethylated, where each methyl state confers different meaning from a biological standpoint (Zhang and Reinberg, 2001). In epigenetic control of chromatin biology, the high degree of modification complexity and coding potential of histone lysine methylation can be explained by the existence of large number of SET domain proteins (Qian and Zhou, 2006).

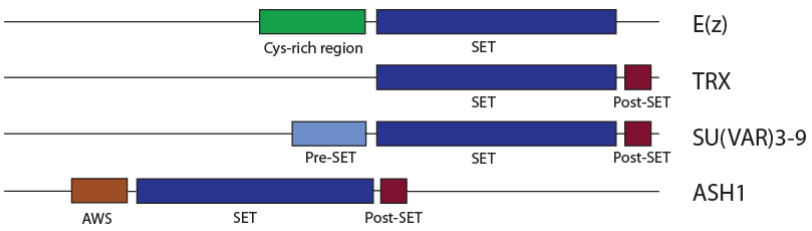

Figure 4. General domain architecture of different classes of SET domain proteins.

Most of the SET domain proteins can be assigned to four evolutionary conserved classes based on the sequence similarity with SET domains found in *Drosophila* as shown in Figure 4. E(z) class proteins have a region with 16-18 cysteine residues in front of the C-terminal SET domain. These proteins are involved in the maintenance of a transcriptionally repressive state of genes via H3K27 tri-methylation. The TRX family does not contain a cysteine rich region in the N-terminal but contain a Post-SET domain consisting of three cysteine residues in the C-terminal that are essential for HMTase activity, and contribute to an active transcriptional state via H3K4me (Zhang et al., 2002). The SU(VAR)3–9 proteins have Post-SET domain like TRX and a Pre-SET domain in front of the SET domain, which are implicated in heterochromatinization via H3K9 methylation. The function of Pre-SET domain is structural, holding two long segments of random coils and stabilizing the SET domain. In contrast to the other three classes the SET domain of ASH1 proteins are centrally placed. Their SET domain is preceded by a cysteine rich Associated with SET (AWS) domain and followed by a Post-SET domain. The ASH1 proteins methylate H3K36 and H3K4me3 on active genes. Additional less conserved classes, like class V proteins of *Arabidopsis* (Ng et al., 2007) conferring H3K27me1 activity (Table 2), may not be present in all organisms. Specificities of different SET domain proteins are predicted from their homology to proteins with known activity and proved by *in vitro* HKMTase assay and *in vivo* studies of changes in histone tails on different target loci in mutants by using different epigenetic methods, for example, chromatin immunoprecipitation (ChIP) or genome wide profiling approaches.

1.3. Structures of SET domain proteins

In the last decade, an enormous progress has been made by the understanding of the 3D structure to identify catalytic mechanism of SET domain HKMTases. As SET domain proteins are functionally very important, it is essential to solve and analyze the structure of this class of proteins. Identify the substrate specificity and product specificity of these proteins is another important issue, which provides the mechanisms of how these proteins function as HKMTases. Structures of the SET domain proteins and its co-domains will give an extensive knowledge of their mode of action, bio-recognition and response process, interaction with target molecules and interacting partners. Some SET domain proteins also catalyze lysine methylation of cellular proteins including cytochrome c, Rubisco, p53 and Taf10 (reviewed in Qian and Zhou (2006)). A number of SET domain proteins structure have been solved to date, and these include ASH1L, NSD1, SMYD2, SMYD3, SET7/9, Dim-5,

Clr4, SET8/PreSET7 and vSET proteins from human, *Neurospora crassa*, yeast and *Paramecium bursaria chlorella virus* 1 (Qian and Zhou, 2006; An et al., 2011; Foreman et al., 2011; Morishita and di Luccio, 2011; Xu et al., 2011). The structures of these proteins reveal that the conserved SET domain has a unique knot-like structure surrounded by a series of β strands. The unusual pseudo-knot is formed by the C-terminal segment of the SET domain consisting of ELxF/YDY and NHS/CxxPN (where x is any amino acid) conserved motifs. A recent review bySchapira (2011) on the structural chemistry of human SET domain proteins described that the domains surrounding the SET domain including Pre-SET and Post-SET are acting as binding interfaces to other proteins or DNA. Different combinations of domains with diverse sequence, structure and electrostatics, would dress the core SET fold in very distinct ways, and allow selective recruitment of interaction partners, or facilitate specific positioning relative to the nucleosome, with functional implications. During the substrate recognition by G9a (Wu et al., 2010), MLL1 (Southall et al., 2009) and SETD7 (Wilson et al., 2002; Kwon et al., 2003; Xiao et al., 2003) proteins, the SET structure remains unchanged, while the conformation of the Post-SET varies considerably. The control of the methylation state by these proteins is effectively controlled by the F/Y switch in the ELxF/YDY where Tyr favors monomethylation and Phe trimethylation (Schapira, 2011).

1.4. SET domain proteins in *Arabidopsis thaliana*

The SET domain proteins of *Arabidopsis* can be grouped in a same evolutionary conserved subclasses as found in *Drosophila* and mammals (Baumbusch et al., 2001; Chen et al., 2010; Thorstensen et al., 2011) and the members of each class are involved in similar processes and methylate the same lysine residue (Kouzarides, 2007). Different classes of SET domain proteins, their HMTase specificity, interaction partners and interacting domains are showed in Table 2.

In animals as well as plants E(z) proteins are part of Polycomb Repressive Complex 2 (PRC2) that suppresses genes by H3K27 trimethylation. Three proteins of the E(z) class have been identified in *Arabidopsis*; CURLY LEAF (CLF), SWINGER (SWN) and MEDEA (MEA) (Baumbusch et al., 2001). During different stages of plant development, *MEA*, *CLF*, *SWN* share at least a subset of common target genes, and during gametophyte and early seed development MEA is required for PcG target repression, whereas during later sporophytic development CLF and SWN take over this function (Makarevich et al., 2006; Wang et al., 2006). Based on genetic, molecular and biochemical evidence, at least three PRC2 complexes

are made and each of them is controlling a particular developmental program. The FERTILIZATION INDEPENDENT SEED (FIS) complex containing MEA, functions during gametophyte development and early seed development in silencing target genes (Komander, 2010). *MEA* and *FERTILIZATION INDEPENDENT SEED2 (FIS2)* are among the imprinted genes in *Arabidopsis*, and are preferentially maternally expressed in the endosperm (Berger and Chaudhury, 2009). CLF and SWN redundantly methylate *PHE1* in sporophytic tissue and low levels of H3K27me3 methylation in *clf swn* double mutants correlate with upregulated *PHE1* expression (Makarevich et al., 2006). The EMF2-PRC2 complex contains CLF/SWN and suppresses premature transition from the vegetative to reproductive stage and takes part in regulating floral organs development (Komander, 2010). VRN2-PRC2 complex contains CLF/SWN and regulates flowering time mediated by vernalization (Komander, 2010).

In the *Arabidopsis* ASH1 class, there are four ASH1 HOMOLOGs (ASHH) and three ASH1 RELATED (ASHR) members was identified (Baumbusch et al., 2001). The functions of most of the proteins in this group have not been investigated yet. ASHH1 can methylate H3 and H4 and *ashh1* mutants are delayed in flowering (Xu et al., 2008; Berr et al., 2009). ASHH2 is considered as a major H3K36me2/me3 histone methyltransferase in *Arabidopsis* (Zhao et al., 2005; Xu et al., 2008). Mutation in *ASHH2* results in a pleiotropic phenotype like small and bushy plants, early flowering, and homeotic changes of floral organs and reduced fertility (Dong et al., 2008; Grini et al., 2009; Ko et al., 2010). The ASH1 RELATED protein ASHR3 is associated with euchromatin and interacts with plant specific transcription factor AMS and the recombinant ASHR3 does not show HMTase activity *in vitro* (Thorstensen et al., 2008). ASHR3 is expressed in the anther tapetum and in the root, regulates the cell cycle and affects stamen development and male fertility (Cartagena et al., 2008; Thorstensen et al., 2008).

The TRX class of SET domain proteins in Arabidopsis consists of two sub-classes. One is *ARABIDOPSIS TRITHORAX* (ATX1-5) and other is *ARABIDOPSIS TRITHORAX RELATED* (ATXR1-7) (Baumbusch et al., 2001). ChIP on different loci suggest that ATX1 and ATX2 confer H3K4 tri- and di-methylation, respectively. There is, however, no genome wide reduction in H3K4me in the *ATX1* or *ATX2* mutants. Flowering time is controlled by a number of pathways that either repress or enhance expression of *FLOWERING LOCUS C*(*FLC*) (Drag et al., 2008; Roudier et al., 2009). All the ATX genes have been tested for changes in flowering time under short day conditions, but only ATRX7 and ATX1 mutants

Class	Name (Arabidopsis)	SDG number (Arabidopsis)	AtGID	HMTase specificity	Method*	Interacts with	Interacting domain	References
I-E(Z)	MEA	SDG5	At1g02580	H3K37me3	ChIP	FIE, FIS2	N-terminal (FIE)	(Makarevich et al., 2006); (Schubert et al., 2006); (Yadegari et al., 2000); (Spillane et al., 2000); (Wang et al., 2006)
	CLF	SDG1	At2g23380	H3K37me3	ChIP	FIE, BLI	CXC(BLI)	(Makarevich et al., 2006); (Schubert et al., 2006); (Schatlowski et al., 2010); (Wang et al., 2006)
	SWN	SDG10	At4g02020	H3K37me3	ChIP	FIE, FIS2, EMF2	C5 (EMF2)	(Makarevich et al., 2006); (Schubert et al., 2006); (Wang et al., 2006); (Chanvivattana et al., 2004)
II-ASH1	ASHH1	SGD26	At1g76710	H3; H4	ON			(Zhao et al., 2005); (Berr et al., 2009)
	ASHH2/ESF	SDG8	At1g77300	H3K4me3; H3K36me2/me3	EH, ChIP, G	H3K4me1,H3 K4me2, H3K4me3	CW	Kim et al. 2005; (Ko et al., 2010); (Zhao et al., 2005); (Grini et al., 2009); (Hoppmann et al., 2011)
	ASHH3	SDG7	At2g44150	–				
	ASHH4	SDG24	At3g59960	–				
	ASHR3	SDG4	At4g30860	H3K4me2; H3K36me2	CH	Aborted micro-spores (AMS)	PHD, SET	(Cartagena et al., 2008);(Thorstensen et al., 2008)
III-TRX	ATX1	SDG27	At2g31650	H3K4me3	P, ChIP	PI5P, WDR5a, TBFII,CLF	PHD/ePHD	(Alvarez-Venegas et al., 2003); (Alvarez-Venegas and Avramova, 2005);(Alvarez-Venegas et al., 2006);(Jiang et al., 2009);(Jiang et al., 2011);(Ding et al., 2011);(Saleh et al., 2007)
	ATX2	SDG30	At1g05830	H3K4me2	ChIP			(Saleh et al., 2008); (Pien et al., 2008)
	ATX3	SDG14	At3g61740	–				
	ATX4	SDG16	At4g27910	–				
	ATX5	SDG29	At5g53430	–				
	ATXR3	SDG2	At4g15180	H3K4me1/me2/ me3	EH, RH, ChIP, G			(Berr et al., 2010); (Guo et al., 2010)
	ATXR7	SDG25	At5g42400	H3K4me1/me2/ me4	ON, ChIP			(Berr et al., 2009); (Tamada et al., 2009)
IV	ATXR5	SDG15	At5g09790	H3K27me1	EH, ChIP	PCNA	PIP Box	(Jacob et al., 2009)
	ATXR6	SDG34	At5g24330	H3K27me1	EH, ChIP	PCNA	PIP Box	(Jacob et al., 2009)
V-SU(VAR) 3-9	SUVH1	SDG32	At5g04940	–				
	SUVH2	SDG3	At2g33290	H3K9me1/me2; H4K20me; H3K27me2	EH	CG methylation	YDG/SRA	(Johnson et al., 2008); (Naumann et al., 2005); (Ay et al., 2009)
	SUVH3	SDG19	At1g73100					
	SUVH4/KYP	SDG33	At5g13960	H3K9me1/me2	P, ChIP, MS	CHG methylation	YDG/SRA	(Wang et al., 2006); Ebbs (Ebbs and Bender, 2006)
	SUVH5	SDG9	At2g35160	H3K9me1/me2	EH, P, ChIP	CG, CHG, CHH methylation	YDG/SRA	(Ebbs and Bender, 2006); (Rajakumara et al., 2011)
	SUVH6	SDG23	At2g22740	H3K9me1/me2	ChIP	CHG, CHH methylation	YDG/SRA	(Johnson et al., 2008);(Wang et al., 2006); (Ebbs and Bender, 2006)
	SUVH7	SDG17	At1g17770	–				
	SUVH8	SDG21	At2g24740	–				
	SUVH9	SDG22	At4g13460	–		CHH methylation	YDG/SRA	(Johnson et al., 2007)
	SUVR1	SDG13	At1g04050	–				
	SUVR2	SDG18	At5g43990	–				
	SUVR3	SDG20	At3g03750	–				
	SUVR4	SDG31	At3g04380	H3K9me2/me3	EH, ChIP, MS	Ubiquitin, H2Bub1	WIYLD	(Thorstensen et al., 2006);.(Veiseth et al., 2011)
	SUVR5	SDG6	At2g23740	H3K9me2, H3K27me2	ChIP	AtSWP1	Not determined	(Krichevsky et al., 2007b; Krichevsky et al., 2007a)

Table 2. SET domain-containing proteins in plants (table modified from Thorstensen et al. (2011)).

flowered more rapidly than the wild type (Tamada et al., 2009). These mutations result in reduced *FLC* transcript levels and reduction in H3K4me3 or H3K36me3 marks on *FLC* chromatin, and an increase in H3K27me3. ATXR7 is associated with the transcription start site of *FLC* chromatin, which has confirmed by the ChIP experiments (Berr et al., 2009).

The SU(VAR)3-9 class of SET domain proteins is involved in silencing of transposable sequences which are highly abundant in centromeric and pericentromeric heterochromatin. This class of proteins consists of 14 different proteins; the SU(VAR)3-9 HOMOLOGs SUVH1-SUVH9 and the SU(VAR) 3-9 RELATED proteins SUVR1-5, which are mainly associated with H3K9 methylation (Baumbusch et al., 2001). Experimental data suggest that H3K9 mono- and dimethylation is carried out by the SUVH proteins and are associated with chromocenters, and involved in gene silencing (Jackson et al., 2004; Ebbs and Bender, 2006; Fischer et al., 2006). SUVH proteins link DNA methylation and the epigenetic gene-silencing marks H3K9me2, and repress transcription of transposons and inverted repeat sequences by directing CHG methylation via CMT3 DNA methyltransferase. SUVH proteins can regulate the expression and chromatin compaction of epialleles, and act together with non-CG methyltransferases and proteins of the siRNA pathway involved in RNA-directed DNA methylation (RdDM) of repeats and transposons (reviewed by Thorstensen et al. (2011)). Three of the SUVR proteins contain a plant specific WIYLD domain and the subnuclear localization of these proteins may be regulated by alternative splicing (Thorstensen et al., 2006). *In vitro* SUVR4 has been found to be a histone H3K9 methyltransferase with a strong preference for monomethylated substrate (Thorstensen et al., 2006). No enzyme activity has been reported for SUVR1 and SUVR2 yet.

1.5. Co-domains in *Arabidopsis* SET domain proteins

In addition to the SET domain, several conserved domains are present in all HKMTases. Co-domains may be involved in histone recognition and responsible for the specificity of particular modified residues (e.g. acetylation or methylation of lysine) in the context of its surrounding amino acid sequence, and for the state of modification (e.g. H3K9me1 vs. H3K9me3) (Taverna et al., 2007). A number of domains in the Royal Super family, including Tudor domain, PWWP domain, MBT domain, chromodomains and PHD finger family recognize and bind methylated lysine residues on the histone tails (Yap and Zhou, 2006). It has been shown that different domains target different histone methylation mark contexts. For example, chromodomains target di- and tri- methyl lysine in H3K9 and H3K27; double

chromodomains of CHD1 target methyllysine in H3K4 context; double Tudor domains in JMJD2A target H3K4me3; tandem Tudor domains of 53BP1 target mono- and di- H3K20me, BPTF, Yng1p and ING2 PhD fingers target di- and tri- K3K4me; PHD finger of BHC80 and cysteine rich ADD domain of DNMT3L target lysine in H3K4 (reviewed in Taverna et al. (2007)).

A number of conserved co-domains of *Arabidopsis thaliana* are showed in Figure 5. In E(z) orthologs, several domains have been identified, including the DNA binding SANT domain, (SWI3, ADA2, N-CoR and TFIIIB), two E(Z) specific domains, EZD1 and EZD2 and a cysteine-rich CXC domain (reviewed in Thorstensen et al. (2011)). The function of the N-terminal charged amino acids-rich EZD1 domain is presently unknown. However, binding to the *Arabidopsis* PcG protein EMBRYONIC FLOWER2 (EMF2) is mediated by the EZD2 domain, also called C5 due to the presence of five signature cysteines. Moreover, for the HKMTase activity of E(z) in *C. elegans* orthologs the cysteine rich CXC domain is required (reviewed in Thorstensen et al. (2011)).

The CW domain is found in *Arabidopsis* ASHH2, which is named after its conserved cysteine (C) and tryptophan (W) residues comprising about 60 amino acids. The CW domain was first identified as an MBD-associated domain (MAD) in a subgroup of methyl-CpG-binding proteins of *Arabidopsis* (Berg et al., 2003). CW domains in human, vertebrates, parasites and higher plants have also been identified (Berg et al., 2003; Perry and Zhao, 2003). Some of the genes that encode CW domain proteins have mutant alleles with phenotypes that underscore their functional importance. Mutation in the mouse *Morc1* causes arrested spermatogenesis, *Morc2b* is involved in hybrid sterility, and *MORC4* has been found highly expressed in large B-cell lymphomas (Inoue et al., 1999; Liggins et al., 2007; Mihola et al., 2009). During the vegetative growth, the *Arabidopsis val1val2* double mutant fails to repress embryonic development (Suzuki et al., 2007). The mammalian CW protein AOF1/LSD2 is a H3K4me1- and me2- specific histone demethylase and has a demethylase-independent repressor function, which, on the other hand, requires the CW domain (Karytinos et al., 2009; Yang et al., 2010).

PHD finger domains are found both is the ASH1 and Trithorax class of SET domain protein. This domain has been shown to interact with di- and tri- methylated H3K4 (reviewed in Taverna et al. (2007)). Human BPTF contains a PHD finger, shown to be a reader of H3K4me2 and H3K4me3 (Wysocka et al., 2006). H3K4me3 can be also read by the human INHIBITOR OF GROWTH2 (ING2) protein (Pena et al., 2006; Shi et al., 2006). So far no

PHD fingers found in SET domain proteins that have reported to bind H3K4me3. But, the PHD fingers of both ATXR5 and ATXR6 strongly bind unmethylated H3 tail peptides (amino acids 1–21), and this binding is negatively affected by the presence of H3K4 methylation (Jacob et al., 2010).

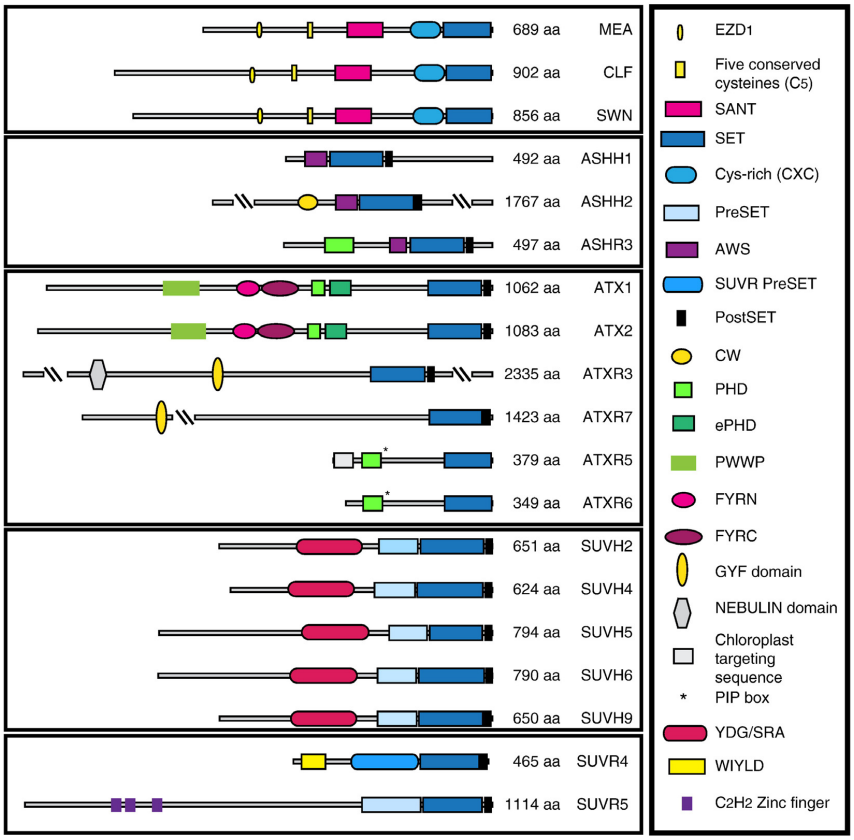

Figure 5. Domain structure of *Arabidopsis* SET domain proteins involved in development (figure form Thorstensen et al. (2011)).

PWPP, FYRN, FYRC, GYF domains are found in the Trithorax class of SET domain proteins. A yeast PWWP domain has been shown to bind H4K20me and to be required for the localization and HKMTase activity of Set9 on this residue (Wang et al., 2009). FYRN and FYRC domains are associated with PHD, SET and PWWP domain with an unknown function. The more conserved parts of these domains are called ATA1 and ATA2 in human ALR protein (Prasad et al., 1997) and in FYR in plant proteins (Balciunas and Ronne, 2000). The GYF (glycine-tyrosine-phenylalanine) domain is known as a member of the super family

of recognition domains for proline-rich sequences (Kofler and Freund, 2006) found in ATXR3 and ATXR7 proteins. When associated with SET domain proteins, the function of this domain is still unknown.

In the SU(VAR)3-9 class of SET domain proteins, several domains are found in the N-terminal part of the proteins including YDG (named for three conserved amino acids)/SRA (SET and RING associate) domain, WIYLD domain and C_2H_2 Zinc finger domains. The YDG motifs are found in all the SUVH proteins, and in plants, heterochromatin is associated with H3K9me1/me2 and controlled by four of the SUVH proteins that through their YDG/SRA domain N-terminal to the SET domain recruit DNA methylation that reinforces the silent state (Jackson et al., 2002; Naumann et al., 2005; Ebbs and Bender, 2006). YDG/SRA domains of KYP/SUVH4 and SUVH6 preferentially bind methylated CHG DNA, suggesting a role of DNA methylation in recruiting H3K9 methyltransferases (Johnson et al., 2007). The plant specific WIYLD domain is found in the SUVR1, SUVR2 and SUVR4 proteins (Thorstensen et al., 2006).

2. Challenges

The complex structure of chromatin contains a number of post-translational modification on the N-terminal histone tails protruding from nucleosomes which are creating binding sites for chromatin-associated sensor and effector proteins facilitating or restricting transcription. The histone lysine methyltransferases (HKMTases) of model plant *Arabidopsis thaliana* consist of a big group of around 40 different SET domain proteins. These SET domain proteins also contain co-domains, some of which are also histone recognition modules. Different HMTase recognize different substrates and the product specificities are also different from protein to protein. The biological functions and targets for these proteins are also divergent. Thus, it is an incredibly huge challenge to understand all the function of these SET domain proteins and point out the substrate and product specificities. In the last decade, a massive progress has been done in this field, but still it is not sufficient for a comprehensive understanding of the possible mechanisms and functions of these proteins. Further *in vitro*, *in vivo* experiments, together with structural analysis of these proteins and their co-domains will provide valuable information about how these protein functions as HKMTases.

3. Aim of study

The overall goal of this thesis has been to broaden our knowledge about the structure and novel functions of the different co-domains present in the SET domain proteins. HKMTase activity has been demonstrated for the *Arabidopsis* SUVR4 and ASHH2 proteins prior to this study. SUVR4 was shown to be responsible for methylation of H3K9 and ASHH2 is considered to be the major H3K36me2/me3 HKMTase (Zhao et al., 2005; Thorstensen et al., 2006; Xu et al., 2008). Co-domains which are thought to be important for the functions of these HKMTases have been identified including the WIYLD domain in SUVR proteins and the CW domain in ASHH2. The WIYLD domain is a plant specific N-terminal domain of SUVR1, SUVR2 and SUVR4 and the cysteine rich CW domain found in a small number of chromatin-related proteins in animals and plants.

The main objectives have been to identify:

- biological and biochemical functions of selected *Arabidopsis* SET domain proteins
- novel functions of co-domains in selected *Arabidopsis* SET domain proteins
- the three-dimensional structure of selected co-domains in *Arabidopsis* SET domain proteins.

4. Results and Discussion

4.1. The SUVR4 and ASHH2 SET domain proteins have different biological functions

Although different HKMTases contains the same conserved SET domain, the high number of SET domain proteins and their diverse expression patterns may reflect a high complexity of epigenetic control and activity during plant development. As described above, a number of *Arabidopsis* SET domain proteins divided into several evolutionary conserved classes according to their sequences and domain architectures are found: i) The E(z) class involved in the maintenance of a transcriptionally repressive state of genes, ii) SU(VAR)3-9 proteins implicated in heterochromatinization, iii) TRX family maintains transcriptionally permissive chromatin states of genes and iv) ASH1 proteins associated with transcriptional elongation. One protein from the SU(VAR)3-9 class (SUVR4) assumed to be involved with heterochromatin and for which no morphological phenotype was found, together with another protein from ASH1 class (ASHH2) with previously described pleiotropic mutant phenotypes were chosen for this thesis.

4.1.1. SUVR4 is involved in repression of transcription of transposons

The SUVR proteins differ from the SUVH proteins in their domain structure, and three closely related SUVR proteins contain a novel WIYLD domain at their N-terminus, and a SUVR specific region preceding the SET domain (Thorstensen et al., 2006). The WIYLD domain consists of about 60 amino acids, only found in plant proteins (Figure 6 and 7).

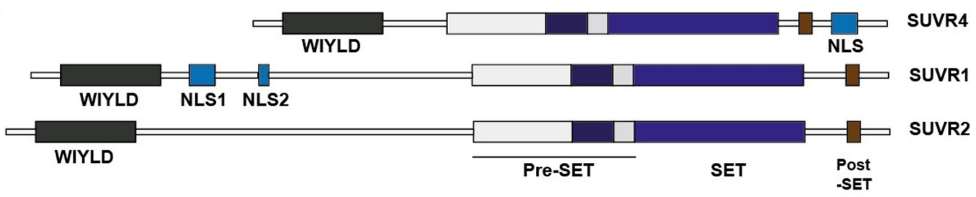

Figure 6. Architecture of SUVR proteins. All the SUVR proteins contain conserved SET domains with a pre-SET and post-SET and a conserved WIYLD domain in the N-terminus end. Nuclear localization signals (NLS) have also been identified (Thorstensen et al., 2006).

All the *SUVR* genes show strongest expression in inflorescences, weakest in the leaves and relatively weaker in the roots than in seedling. Its ubiquitous expression pattern suggests that *SUVR* genes are important during the whole life cycle of the plant (Thorstensen et al., 2006). All the *SUVR1*, *SUVR2* and *SUVR4* genes express alternative spliced mRNAs, which may regulate the subnuclear localization. SUVR4-GFP fusion proteins expressed with a glucocorticoid-inducible construct were localized in foci of unknown function and to the nucleolus, and suggested that these proteins might be involved in regulation of rRNA expression (Thorstensen et al., 2006). We also generated GFP overexpression (OE) lines where SUVR4-GFP expression was driven by the strong, constitutive Cauliflower mosaic virus 35S promoter, giving a uniform SUVR4-distribution in the nucleus in addition to accumulation in the nucleolus. This opened for new hypothesis on function, and was shown (Paper I) that SUVR4 is involved in repression of transposons.

In this thesis the mechanisms by which SUVR4 represses transposon activity has been studied. Expression analysis of the transposons *MULE At2g15810*, *AtIS112A* (At4g04293) and *AtCOPIA4* in the SUVR4 overexpression line, showed significant reduction in expression for all studied transposon, compared to wild type line, suggesting that SUVR4 acts as a repressor of transposable elements. The transcription level of *MULE* was induced 2.5-3 folds in a SUVR4 RNAi line compared to wild type line. Transposon *AtS112A,* with an intermediate expression level was only affected in the OE line. The variable release of repression in the RNAi line suggested that SUVR4 regulates transposon activity in a locus specific manner, where activity of SUVR4 alone was sufficient for repression of *MULE*. However it works redundantly with an unknown HKMTase at other elements like *AtS112A*, *AtMU1* and *AtSN1*.

4.1.2. ASHH2 is needed for normal plant development

ASH1 HOMOLOG2 (ASHH2), also known as EFS/SDG8, was first characterized as a repressor of transition from vegetative to reproductive growth, which is accompanied by the down-regulation of the flowering repressor gene *FLC* (Schotta et al., 2002; Zhao et al., 2005; Xu et al., 2008). *ashh2* mutants display early flowering correlating with reduced *FLC* transcript levels. Mutation in *ASHH2* has also been shown to affect the expression of two key regulators of shoot branching, *SPS/BUS* and *UGT74E2*, resulting in increased shoot branching and small, bushy plants (Dong et al., 2008). The *carotenoid chloroplast regulatory1 (ccr1)* mutant identified based on altered carotenoid composition, also turned out to be an *ashh2*

mutant (Cazzonelli et al., 2009). *ASHH2* is required for expression of CAROTENOID ISOMERASE (*CRTISO*) and downregulation of *CRTISO* in the *ccr1* mutant (Cazzonelli et al., 2009). Carotenoid pigments are critical for plant survival, and the altered mutant carotenoid profile may partially affect shoot branching. *ASHH2* is furthermore involved in the regulation of organ development in the flower, it has a role in ovule and anther development, and due to developmental defects of reproductive organs *ashh2* mutants have a very low seed set (Grini et al., 2009). Recent publications showed that *ASHH2* additionally is required for basal and R protein mediated pathogen resistance in *Arabidopsis* and play a crucial role in plant defense against fungal pathogens by regulating a subset of genes within the jasmonic acid (JA) and/or ethylene signaling pathway (Berr et al., 2010; Roudier et al., 2011). So, ASHH2 affects flowering time, branching, reproductive organ and pathogen defense and the pleiotropic phenotype conferred by the mutations in *ashh2* suggest that ASHH2 controls many key regulatory genes, and therefore has general fundamental importance for plant development.

4.2. Biochemical functions of SUVR4 and ASHH2

In general, different HKMTases recognize different lysine residues on histone tail as substrate and responsible for methylation. Current research suggested that functions of the SET domain proteins are not dependent only on the specificities of the SET domain; it is also dependent on the co-domains of these proteins. Several conserved domains are present in addition to the SET domains in the *Arabidopsis* SET domain proteins, which may contribute to the recruitment of the histone modifiers to relevant sites in chromatin (Ruthenburg et al., 2007) or to modulate the activity of the methyltransferases. The activity of the SET domain proteins is also influenced by other interacting proteins, which can change the product specificity of the HKMTases as well. In this study, I have elucidated how the co-domains, WIYLD of SUVR4 and CW of ASHH2, can influence or regulate the biochemical function of SUVR4 and ASHH2 proteins.

4.2.1. SUVR4 activity and role of ubiquitin binding to WIYLD domain

Immunocytological analysis on seedling leaves showed a strong reduction in H3K9me1 and corresponding increase in H3K9me3 in nuclei with high SUVR4-GFP expression (Paper I). However, low SUVR4-GFP expressed nuclei did not show this effect, suggesting that the global changes in H3K9me1 and H3K9me3 correlated with SUVR4-GFP expression. It was previously shown that SUVR4 methylate calf thymus histone H3, but not the recombinant

full-length H3 (Thorstensen et al., 2006). This indicated that post translational modification of H3 is necessary for SUVR4 activity, and discovered that SUVR4 specifically methylate H3K9 with a preference for monomethylated H3K9. It was suggested that the WIYLD domain might be involved in directing proteins to their targets, or conversely be directed to its targets through interactions with the WIYLD domain (Thorstensen et al., 2006).

To understand the role of the WIYLD domain and thereby elucidate SUVR4 function, this domain was cloned and used for functional analysis *in vivo* and *in vitro* (Paper I and II). Yeast two-hybrid (Y2H) screening showed that the WIYLD domains of SUVR proteins were interacting with the UBIQUITIN EXTENTION PROTEIN 1 (UBQ1, At3G52590) which consists of ubiquitin and the ribosomal protein L40. Y2H assays demonstrated that the interaction is only between the N-terminal ubiquitin of UBQ1 with WIYLD, not the L40, suggesting that the WIYLD domain specifically binds the ubiquitin. In *Arabidopsis* H2B is monoubiquitinated on lysine 143 (H2Bub1) (Sridhar et al., 2007). A pull-down assay showed that S4WIYLD also binds H2Bub1 (Paper I), and ELISA provided confirmation of binding to free ubiquitin (Paper II). The WIYLD domains of SUVR1 and 2 have high sequence similarity to that of SUVR4 (Figure 7), and Y2H and ELISA established that the WIYLD domain of SUVR1 and 2 also can bind ubiquitin (Paper II). In ELISA, the binding affinity was slightly higher for S2WIYLD and somewhat weaker for S1WIYLD relative to S4WIYLD. Thus, together the Y2H, pull-down and ELISA indicated that all WIYLD domains of SUVR proteins bind ubiquitin.

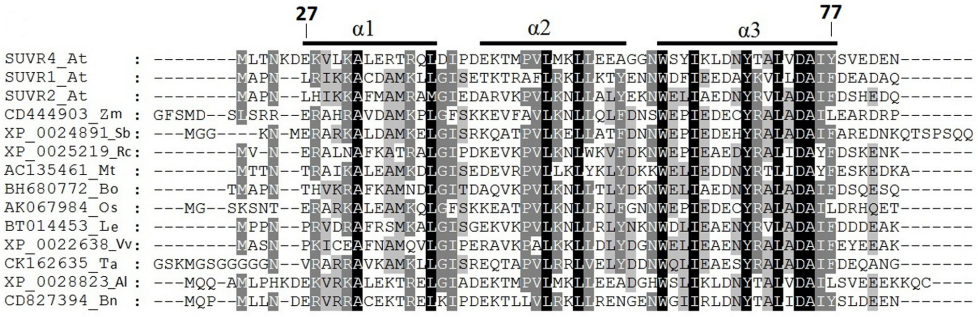

Figure 7. Conserved residues and multiple sequence alignment of SUVR4 WIYLD domain named after its conserved C-terminal residues. This is a leucine rich domain and in the C-terminal part of this domain has 5 conserved (Trp (W), Ile (I), Tyr(Y), Leu (L) and Asp (D)) amino acids. According to PhDsec and JPRED secondary structure prediction, this domain contains three alpha helices (Thorstensen et al., 2006).

The crucial question was how the ubiquitin-binding of WIYLD would affect SUVR4 function. Therefore SUVR4 HKMTase activity was tested with and without the presence of the WIYLD domain. The SUVR4 full-length protein showed higher enzyme activity than the protein without the WIYLD domain, suggesting that the WIYLD domain has a positive effect on enzyme activity of SUVR4 although the WIYLD domain itself does not have any HKMTases activity (Paper I).

Ubiquitin binding proteins are important for regulating the stability, function and/or localization of ubiquitinated proteins. Several enzymes that are involved in ubiquitin pathways have shown to be regulated by ubiquitin. A deubiquitinating enzyme ataxin-3, a polyglutamine disease protein is directly regulated by the ubiquitination (Todi et al., 2009). Ubiquitin binding is crucial for the optimal catalytic activation of deubiquitinating enzyme isopeptidase T (IsoT, or USP5) (Reyes-Turcu et al., 2006). Thus we tested whether the ubiquitin could stimulate the SUVR4 HKMTase activity. Addition of ubiquitin in the enzymatic reaction increased the activity of SUVR4 indeed. In the presence of free ubiquitin the full-length protein was stimulated by 2-3 folds whereas the SUVR4 construct without the WIYLD domain was only weakly affected, suggesting that most of the ubiquitin response mediated through the WIYLD domain (Paper I).

SUVR4 protein has very strict substrate specificity because no histones other than H3 were methylated by SUVR4, even after adding free ubiquitin to the enzymatic reaction. However, the addition of free ubiquitin converted the protein from a strict dimethylase to a di/trimethylase *in vitro* (Paper I) and this was confirmed by the peptide mass fingerprinting experiments. Therefore an important question has been whether SUVR4 acts the same way *in vivo*. Over-expression of SUVR4 showed a massive shift from H3K9me1 to H3K9me3 *in vivo*, suggesting that ubiquitin either in its free form or conjugated to other proteins like H2B could act as signal for H3K9 methylation. These *in vitro* and *in vivo* data suggest that to efficiently convert H3K9me1 to H3K9me3 *in vitro* another component is needed in addition to ubiquitin for SUVR4.

4.2.2. ASHH2 has a preference for H3K4 actively transcribed genes

Chromatin of *ashh2* mutants show a global reduction on H3K36me2/me3 and therefore ASHH2 is considered to be the major enzyme for H3K36me2/me3 in *Arabidopsis*. To identify the features of the chromatic context in which ASHH2 is acting, experiments were carried out

to investigate the effect of *ashh2* mutation on expression and histone marks for a selected panel of genes (Paper III). ChIP analysis was done on a set of tissue-specific genes with differential expression profiles in seedling and flowers downregulated in the *ashh2 mutant* and non-affected genes with high expression, to compare wild type (wt) and *ashh2* mutant seedlings with antibodies against H3K4me3, H3K36me2 and H3K36me3. The tissue specific genes tested showed very low H3K4me3 level and this mark was largely unaffected by the *ashh2* mutation. Strongly expressed genes showed a high H3K4me3 level in the wt and low level in the mutant although the transcript level was not affected. H3K36me2/me3 levels were low in seedlings for inflorescence transcription factor genes and unaffected by *ashh2* mutation. The *ashh2* mutant showed significant increase of K36me2 and reduction of K36me3 or reduction in both marks for seedling and constitutively expressed genes. These data suggested that the level of expression of a gene might be reflected by the level of H3K36me3 methylation, but that this mark was not required for expression of genes with high expression level. However, H3K36 trimethylation by ASHH2 is positively correlated with transcription of tissue specific genes.

To investigate whether the genes with *ASHH2* dependent regulation had particular characteristics, an analysis of the presence of H3K4me3, H3K27me3 and H3K36me2 using published global ChIP data (Oh et al., 2008; Zhang et al., 2009), was done for the genes downregulated in *ashh2* mutant plants according to microarray experiments (Paper III) (Xu et al., 2008; Cazzonelli et al., 2009). Over 84% of the genes downregulated in the mutant had H3K4me3 marks; often in combination with H3K4me1 and me2, suggesting that ASHH2 associates with transcribed genes. Genes likely to be silent with H3K27me3 marks were underrepresented and tissue specific or developmentally regulated genes with H3K4me3, H3K27me3 and H3K36me2 as well as genes encoding transcription factors were overrepresented.

Another survey was done for H3K4me1, me2, me3 and H3K27me3 using a published, global data set for *Arabidopsis* seedlings (Zhang et al., 2009) with the 45 downregulated genes in *ashh2* mutants and the genes used in the ChIP experiments in Paper III. The inflorescence-specific genes were only marked by the repressive K27me3, highly expressed genes marked K4me1/me2/me3 or K4me1/me3, and tissue-specific seedling expressed gene like *FLC* marked both K4me2/me3 and K27me3 and *AtDMC1* marked with K4me1/me2 and *MAF1* marked with K4me1. Genes devoid of H3K4me marks were significantly underrepresented

among 45 *ASHH2*-dependent genes compared to global data (Zhang et al., 2009). The genes with K4me1 marks were similar to wt (31.1 versus 32%), while K4me2 and K4me3 were overrepresented. Combinations of K4me marks, K4me2/me3 and K4me1/me2/me3 were overrepresented among the genes downregulated in *ashh2* mutant seedlings. These results suggested that ASHH2 is associated with actively transcribed genes and has a particular preference for transcribed genes with K4me2 and K4me1/me2 marks.

Recent *in vitro* experiments identified that ASHH2 can confer methylation of both H3K36 and H3K4 and this HKMTase activity seem to increase in the presence of *FLC* activator protein FRIGIDA (FRI) (Ko et al., 2010). The experiments were conducted in the Col ecotype which is mutant for *FRI,* and in this genetic background there is little that indicates that ASHH2 confers H3K4me3 activity. The reduction of H3K4me3 and H3K36me3 for the highly expressed *ACTIN2* and *GAPA* genes was not accompanied by reduced expression levels. Changes in H3K4me3 or in H3K36me2 were not detected for the tissue-specific genes expressed in mutant inflorescences. All seedling-expressed genes had lower levels of H3K36me3 in the mutant, but two genes showed an increase in H3K36me2 methylation. This may suggest that *Arabidopsis* has another SET-domain protein that is responsible for H3K36me2, which is used by ASHH2 as a substrate. Together, these results support that the major activity of ASHH2 is H3K36 tri-methylation in *fri* background.

4.2.3. CW domain is a new histone recognition module

The ASHH2 protein is one of the largest of the SET domain proteins of *Arabidopsis* and the CW domain was identified in addition to the SET domain. The CW domain is a cysteine rich domain found in a small number of chromatin-related proteins in animals and plants shown in the multiple sequence alignment (Figure 8). In this thesis the function of CW in relation to ASHH2's activity was investigated. Since ASHH2 seemed to have a preference for H3K4me marked genes, it was tested whether CW could be a reader of such marks. We have expressed and purified the ASHH2 CW domain as a GST fusion protein and tested it to a panel of immobilized histone tail peptides. The pull-down assay showed that the ASHH2 CW domain binds and shows preference for mono- and di-methylated H3K4 peptides (Paper III). This was confirmed by surface plasmon resonance data that showed that the affinity for the mono- and di-methylated peptides is in the micromolar range which is comparable to PHD fingers and other histone recognition modules.

A pull-down assay was done with chromatin prepared from *Arabidopsis* seedlings and the ASHH2-CW protein to check if the CW domain can bind histone H3K4me tails in a nucleosomal context (Paper III). The CW domain protein pulled down histones that were mono-, di-, or tri-methylated at H3K4 in contrast to a mutant version of the CW domain (W874A), suggesting that the binding of CW to H3K4me is specific. This indicates that the CW domain can bind histone H3K4me tails in a nucleosomal context.

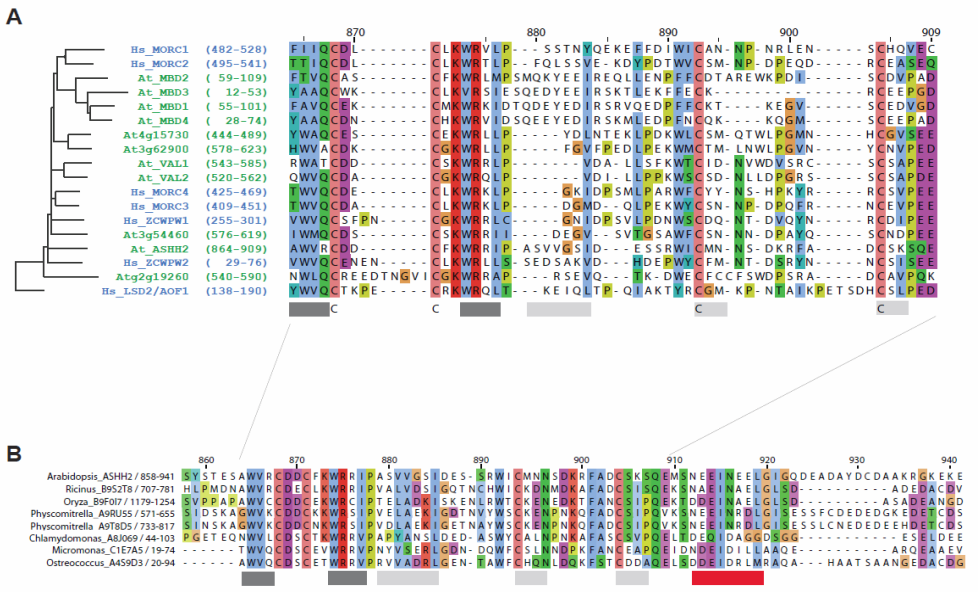

Figure 8. Multiple sequence alignments of (A) human and plant CW domains and (B) selected plant orthologs of ASHH2 CW domains.

As the ASHH2 CW domain binds H3K4me1 and me2, this can consider in more details the functional implication of this reading module in a context with its role as an HKMTase. ASHH2 is considered to be the major enzyme for H3K36me3 in *Arabidopsis* in *fri* background. Thus, ASHH2 seems to be a protein with an H3K4me1/me2 reading module and an H3K36me3 writing module.

4.3. Targeting of SET domains

Different modifications of lysine residues by SET domain proteins relate to different chromatic distribution and functions. Histones mono-, di-, or trimethylated at lysines are differently distributed within eu- and heterochromatin, each potentially indexing a specific

biological outcome. Modifications can affect one another by crosstalk; hence one specific modification can repress or facilitate other modifications on nearby residues, possibly leading to an amplification of the chromatin structure alteration. Another consequence is the subsequent altering of recruitment of proteins recognizing different modification patterns.

4.3.1. Cross-talk determines the activity of SUVR4

SUVR4 preferably uses H3K9me1 as substrate and H3K9me1/2 is distributed in chromocenters and pericentric heterochromatin, whereas in *Arabidopsis* H3K9me3 methylation broadly marks euchromatin and a detectable level found in regions with silenced transposon and pseudogenes (Naumann et al., 2005; Roudier et al., 2011).ChIP experiments confirmed that SUVR4 is associated with genes both in eu- and heterochromatin, but a significantly higher amount of SUVR4-GFP found at euchromatic genes. However, only transposon and pseudogenes were affected by overexpression of SUVR4, showing an increase of H3K9me3 and reduction of H3K9me1 (Paper I). Another ChIP analysis of two of the transposons in knock-down SUVR4 RNAi line showed an increase of H3K9me1 on transposons and a corresponding reduction of H3K9me3. Both *in vitro* and *in vivo* data confirmed that SUVR4 has no HKMTase activity on selected euchromatic genes, but specifically targets transposon and repeat sequences where it converts H3K9me1 to H3K9me3. No activity of SUVR4 on euchromatic genes could be related to the cross-talk to PTMs characteristic for euchromatin. HKMTase activity of SUVR4 was not affected by monomethyl H3K4, whereas trimethyl H3K4 reduced the activity, arguing that chromatin of euchromatic genes, with a high level of this mark, might not be good substrate for SUVR4 activity.

Pericentric heterochromatin most likely is not the preferred target of SUVR4 activity because of the high level of uninterrupted H3K9me2 (Bernatavichute et al., 2008) and the cell cycle dependent H3S10ph modification generated by Aurora kinase 1 inhibits SUVR4 activity *in vitro* (Demidov et al., 2009). But SUVR4 could potentially add another methyl group on H3K9me2 marked transposons in these regions before H3S10ph is added by Aurora kinase 1 and when ubiquitin levels are high.

A ChIP analysis showed a low level of H2Bub1 was found both in the WT and *ubp26* (ubiquitin-specific protease 26) mutant line at tested transposons. Distribution of the euchromatic mark H2Bub1 was not affected by SUVR4 overexpression at any of the tested

transposon sequences. A reduction of both H3K9me2 and H3K9me3 at transposon was observed in a line (*ubp26-1*) with mutation in the histone H2B deubiquitinase gene *UBP26/SUP32* on the same sequences targeted by SUVR4, suggesting that UBP26/SUP32 acts upstream of SUVR4 in the same pathway, leading to repression of transposon activity. This suggests that the reduction of H3K9me3 in *ubp26-1* mutant background could be due to reduced SUVR4 activity. UBP26 can repress the transposons transcription by lowering the H2Bub1 level to maintain the repressive H3 methylation (Sridhar et al., 2007) and/or by maintaining a high level of free ubiquitin which stimulates the SUVR4-mediated H3K9me3. UBP26 can also cleave the UBQ1 (ubiquitin extension protein) to obtain the free ubiquitin, as it does for CEP52 (Sridhar et al., 2007). However, no reduction of free ubiquitin in the nuclear extract of *ubp26-1* mutants and no effect on H3K9me3 or H2Bub1 at transposon sequences in a line (*hub2-2*), with mutation in the H2B E3 ubiquitin-protein ligase *BRE1-like 2* gene, suggest that levels or subnuclear distribution of free ubiquitin can regulate SUVR4-dependent H3K9me2/3.

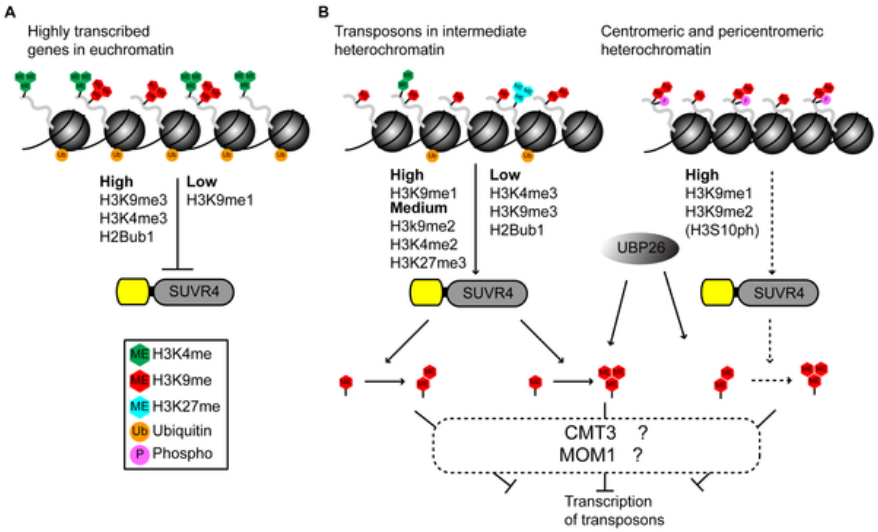

Figure 9. Model describing the relationship between free ubiquitin and SUVR4 activity on transposons.

To describe the activity of SUVR4 on transposons and SUVR4 cross-talk with other PTMs we suggested a model showed in figure 9. SUVR4 has no activity on genes with high H3K4me3, H3K9me3, H2Bub1 and low level of H3K9me1. SUVR4 has a preference for heterochromatic transposons intercalated within euchromatin, maintained by its specificity for H3K9me1, which is highly enriched at transposons and its repression by activating marks like

H3K4me3. The deubiquitinase UBP26 regulates H3K9me2/me3 at the same targets as SUVR4, and might produce free ubiquitin that stimulates the H3K9me2/me3 activity of SUVR4 at target transposons. Although SUVR4 normally is repressed by H3K9me2 and H3S10ph which is high in pericentric heterochromatin, these regions may be targets for SUVR4 activity when ubiquitin levels are high. Since the transposons also contain a medium level of H3K27me3 in addition to H3K9me3, this could possibly create a binding site for CMT3 in order to repress transcription in a DNA methylation-dependent manner at some transposons. At other transposons, transcription may be repressed in a DNA methylation-independent manner by the MOM transcriptional repressor.

4.3.2. CW domain has an important role in ASHH2 functions

ASHH2 targets actively transcribed genes, have a particular preference for transcribed genes with H3K4me2 and H3K4me1/me2 marks, and the CW domain is a histone recognition module with specificity for methylated H3K4. This may indicate that CW may contribute to ASHH2's preference for genes with H3K4 methylation. H3K4me3 is in particular associated with transcribed genes, and H3K4me2 often co-occurs with H3K4me3 in the 5′-end of genic regions, while H3K36me2 increases towards the 3′-end (Oh et al., 2008; Zhang et al., 2009). To test this the DNA was analyzed from seedling chromatin pulled down by the CW domain of ASHH2 followed by real-time PCR to check whether this domain targets genes that are regulated by ASHH2 (Paper III). *FLC* was proven to be targeted by ASHH2 *in vivo* by Ko *et al.* (2010) and was also detected in the chromatin pull-down (ChPD) experiment, suggesting the ability of CW to identify *in vivo* targets of ASHH2. Chromatin associated with the genes that showed substantial reduction in H3K36me3 level in the *ashh2* mutant was strongly pulled down by CW domain, again suggested that CW domain may contribute to the targeting of ASHH2 to chromatin associated with these genes. ChIP experiment on the genes suggested that the recovery profiles are very similar for ChPD and H3K4me1 and H3K36me3. Western blot of ChPD experiments showed that CW efficiently pulled down H3K4me1 marked chromatin of Wt seedling, and antibodies against H3K36me3 revealed the presence of this mark on the chromatin pulled down by CW domain. This suggested that H3K4me1 and H3K36me3 co-reside on the same or neighboring nucleosomes. As expected if the ASHH2 HKMTase activity mediated the H3K36me3 marks, a reduction in H3K36me3 was shown in the chromatin pulled down with the CW domain in the *ashh2* mutant. This substantiated that H3K36me3 mark, mediated by ASHH2 activity, is associated with H3K4me1 which is

preferred target for ASHH2 CW. Therefore, a model was postulated for ASHH2 function and the role of CW domain on ASHH2 function in Figure 10.

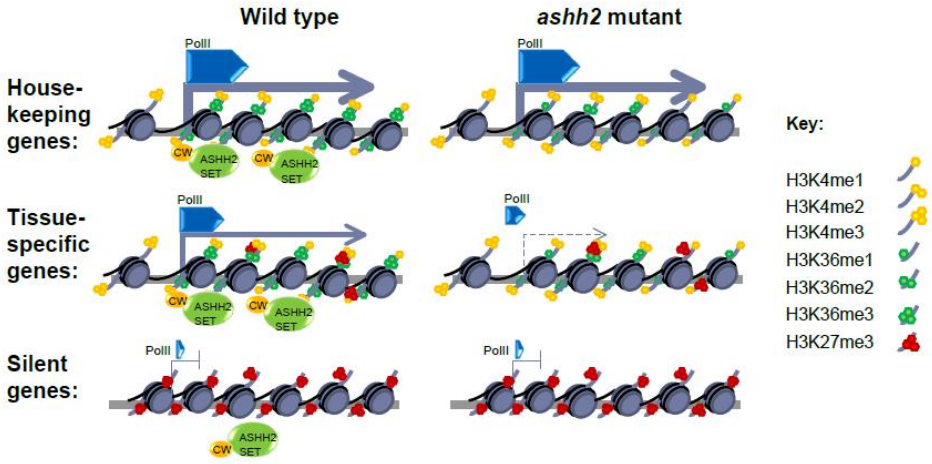

Figure 10. A model for ASHH2 function and the role of its CW domain. Icons for each of the histone modifications are shown to the right. The ASHH2 protein and its SET domain are shown in green, while the CW domain engaged in H3K4me-binding is shown in yellow. RNA polymerase II occupancy is illustrated with blue pentagons, while the extent of transcription is indicated with arrows.

This model described that the CW domain of ASHH2 binds H3K4me2 found close to transcription start site (TSS) in both highly expressed house-keeping genes and weakly expressed tissue-specific genes, but not in silent genes. In addition CW binds H3K4me1 found along the transcribed gene regions. This facilitates H3K36me3 methylation both near the TSS, and along the gene body. Loss of H3K36me3 has little consequences for the transcription level of highly expressed house-keeping genes. However, loss of H3K36me3 due to mutation in *ASHH2* results in non-sustainable transcription of genes that may carry both active and repressive chromatin marks, e.g. H3K4me and H3K27me3, and/or have a low and tissue-specific expression. Due to the lack of H3K4me marks, ASHH2 is not active on silent genes and thus there is no effect on transcription when ASHH2 is mutated.

4.4. Structure sheds light over function

The WIYLD and CW domains are co-domains of *Arabidopsis* SET domain protein that may have similar function: WIYLD domain may read ubiquitin and H2Bub1, and influence the

activity of SUVR4 by the cross-talk between histone marks; whereas CW domain has a preference for mono- and dimethylated H3K4 and contributes to ASHH2 function for maintenance of actively transcribed genes. As WIYLD domain has functional importance for the activity of the HKMTases and CW domain is important for targeting, we went on to identified the important residues responsible for binding affinity and solve the structure of these domains to elucidate the mode of interaction of WIYLD domain to ubiquitin and CW domain to H3K4 peptides.

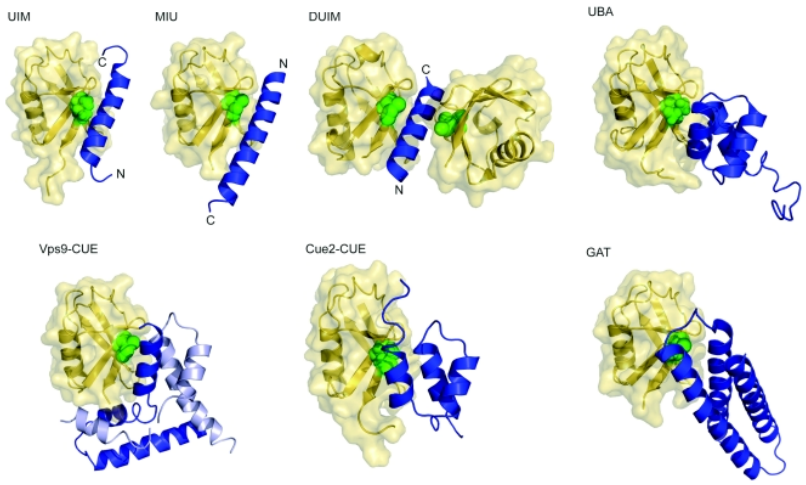

Figure 11. Helical ubiquitin-binding domain (UBD) structures (figure from (Hurley et al., 2006)). The ubiquitin molecule (yellow) in ribbon and surface representations is shown with corresponding helical domain (blue) of UBDs in ribbon representation. Ile44, the centre of the hydrophobic recognition patch on the ubiquitin, is shown as green spheres.

4.4.1. Ubiquitin Binding domains

Ubiquitination is one of the principal PTM of proteins. This PTM helps a cell to control intracellular signaling events and may facilitate conformational changes leading to allosteric regulation or modify interaction surfaces for specific recognition (Garner et al., 2011). Ubiquitin (Ub) binds non-covalently to ubiquitin binding domains (UBDs) found in more than twenty known protein families (Harper and Schulman, 2006; Dikic et al., 2009; Komander, 2009). Covalent binding of ubiquitin to different proteins is regulated by the activity of ubiquitin-activation (E1), ubiquitin-conjugation (E2) and ubiquitin-ligation (E3) enzymes (Dikic et al., 2009). Poly-ubiquitin chains are recognized by the UBDs of receptors that target

proteins for proteosomal degradation, while monoubiquitin is recognized by the UBDs of receptors involved in processes like DNA repair, regulation of protein activity, chromatin remodeling and transcription (Hicke et al., 2005; Dikic et al., 2009; Garner et al., 2011). In general, UBD structures responsible for non-covalent ubiquitin binding can be divided in four categories: 1) α-helices; 2) Zinc fingers (ZnFs); 3) plekstrin homology (PH) domains and 4) Ubiquitin-conjugating (Ubc)-like domains (review in (Hurley, Lee et al. 2006)). These UBDs families differ in structure and the type of ubiquitin modification they recognize. Most of the UBDs use the α-helixes to bind the hydrophobic patch on its β-sheet surface (Garner et al., 2011). Structures of the ubiquitin binding domain-ubiquitin complex have been reported, and which shows that the ubiquitin binding motifs Rabex-5-MIU Vps27-UMI, S5a UIM, Vps9-CUE, Cue2-CUE bind the β-sheet surface region of Ub with their α-helices (Penengo et al., 2006) and (review by (Hurley et al., 2006)), as shown in Figure 11.

4.4.2. WIYLD domain structure identified the binding surface for ubiquitin

A solution structure of the WIYLD domain of the SUVR4 protein (S4WIYLD) has been solved using NMR spectrometry to investigate the mode of interaction (Paper II). PhDsec and JPRED secondary structure prediction suggested that this domain contained three alpha helices as other three-helix bundle UBDs. However, the NMR experiments revealed a helical structure containing four α-helices of the WIYLD domain. Comparison of S4WIYLD structure and UBDs structures suggested that the WIYLD domain shares a common fold like other UBDs, but contains an additional helix in between the second and third helix of common UBDs. It was shown that WIYLD is binding ubiquitin and the conserved Leu39-Asp40-Ile41 motif in turn 1, and the Asn68-Tyr69-Thr70 motif N-terminal to helix 4, are in corresponding positions in WIYLD and UBDs suggests functional significance also for WIYLD interaction with ubiquitin.

The chemical shift perturbation (CSP) method by NMR and mutagenesis analysis substantiated that these two motifs of the C-terminal part of the S4WIYLD domain are involved in ubiquitin interaction (Paper I and II). Alignment of S4WIYLD with other WIYLD domains and UBDs showed that many of the residues are highly conserved. Ubiquitin contains a highly conserved, hydrophobic and concave surface defined by residues L8, I44 and V70, which is usually involved in non-covalent binding of UBDs (Dikic et al., 2009; Garner et al., 2011; Komander, 2009; Seet et al., 2006). Our docking model (Figure 12) suggested that S4WIYLD binds ubiquitin through the exposed hydrophobic surface of its N-

terminal and C-terminal part, including helix 1 (α1), turn 1, turn 3 and helix 4 (α4) to the hydrophobic surface of ubiquitin, as shown for several ubiquitin binding domains including BMSC, p47, Dsk2, and hHR23A (Chang et al., 2006; Mueller and Feigon, 2002; Ohno et al., 2005; Yuan et al., 2004). More precisely, turn 3 and helix α4 of the WIYLD domain are crucial for ubiquitin binding. R37 of α1 and D74 of α4 were identified as important residues, as the mutation of these residues reduced the interaction between H2Bub1/ubiquitin and S4WIYLD domain in *in vitro* pull-down assay and ELISA (paper I and II).

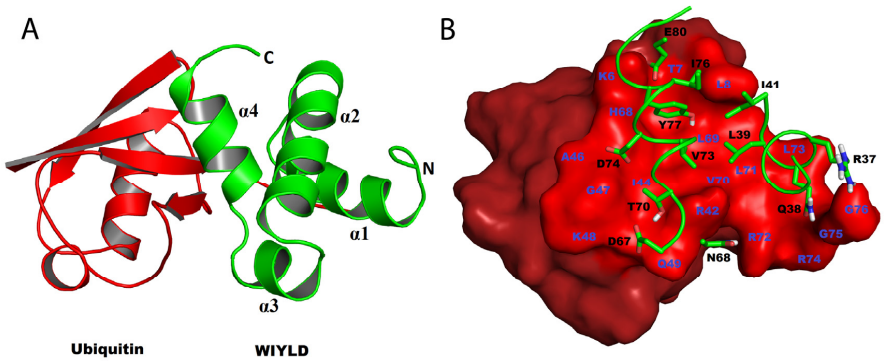

Figure 12. Docking model of SUVR4-WIYLD with ubiquitin. A) Ribbon representation of WIYLD-Ub complex, WIYLD domain in green and Ub in red. B) Representation of the interaction at the WIYLD-Ub interface. Shown is the interaction surface between ubiquitin (surface; firebrick and red) and WIYLD domain (ribbon and stick; green). Ubiquitin residues at the molecular interface are shown in red (residues numbers are in indigo). The side chains of the WIYLD residues (residues numbers are in black) responsible for interacting Ub are shown in stick representation.

The docking model indicated that Arg37 of S4WIYLD binds Gly76 of free ubiquitin. A moderate reduction (29 %) in ubiquitin binding resulted from the R37A mutation. However, this mutation had a dramatic effect regarding interaction with H2Bub1 – a total loss of binding in a pull-down assay (Veiseth et al., 2011). Interestingly, the C-terminal residue of ubiquitin, Gly76, is involved in covalent conjugation of ubiquitin and histone H2B lysine residue (Thorne et al., 1987). Recognition of the C-terminus of ubiquitin is common for proteins regulating ubiquitin conjugation and deubiquitination (review in Reyes-Turcu et al. (2006); Drag et al. (2008); Winget and Mayor (2010)). A reader of ubiquitinated H2B is also likely to recognize the link between ubiquitin and H2B, hence implying that the WIYLD domains of SUVR proteins might be the first readers of H2Bub1 identified to date.

4.4.3. H3K4me-recognizing readers

The CW domain of ASHH2 was identified as a new novel reader of the histone code (Paper III). Histone tails peptide binding was also showed for three other CW domains, VAL1, ZCWPW1 and MORC4, suggested that this is the generic molecular function of CW domains. Unlike the ING2 PHD finger, CW domain of ASHH2 showed preference for mono- and dimethylated H3K4 peptides, while its binding to H3K4me3 is close to background level. VAL1 shows preference for me2 and me3, ZCWPW1 for me3 and MORC4 for me2, suggesting that CW has a novel profile of ligand selectivity among the H3K4-specific recognition modules families. The MBT domain binds several mono- and di-methylated lysines on both H3 and H4 peptides, and shows little sequences selectivity compared to the CW domain. The differences between different H3K4me recognizing readers lead us to solve the structure of ASHH2 CW domain to identify the mode of interaction with histones.

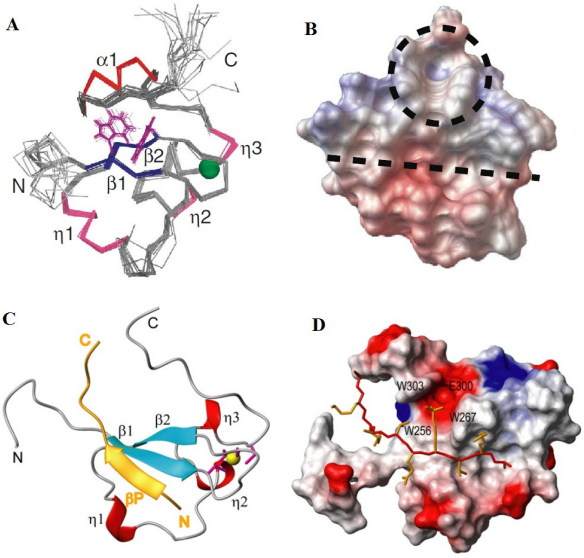

Figure 13. Solution structure of the CW domain and prediction of its binding site for the histone tail. A) Backbone traces of 20 conformers of the solution structure of the ASHH2 CW domain showing residues 858–928. B) Surface representations and surface potential of the ASHH2 CW domain. The predicted placement of the histone tail backbone and the aromatic cage is indicated by a dotted line and a circle, respectively. C) Lowest energy Structure of ZCWPW1 CW, in a ribbon representation. D) Electrostatic surface potential of the ZCWPW1 CW domain in complex with H3(1-10)K4me3 (C and D modified from He et al. (2010)).

4.4.4. Mode of interaction of CW domain with histones

To investigate the mode of interaction of CW domain with the histone tail peptide, the solution structure of the CW domain of ASHH2 protein was solved using NMR spectrometry (Paper III). The solution structures of ASHH2 CW, and of ZCWPW1 CW solved by He et al. (2010), show a common structural core built around two β-strands, a Zn^{2+} - binding sites and three short helical elements (Figure 13). These domains share a similar histone tail binding site containing a cleft traversing one side of the domain, just underneath a pocket containing two conserved tryptophan residues forming an aromatic cage. Figure 13 represents the structures of CW domains and placement of the histone tail backbone.

A chemical shift perturbation experiment identified the residues responsible for interaction of the CW domain and histone peptide. Point mutations of the residues in and around the putative histone tail binding sites confirmed the binding site of the ASHH2 CW domain. Mutation of the three tryptophan (865, 874 and 891) to alanine, abolished the binding of histone tails. First two tryptophans are located in the predicted methyllysine-binding site and the other in the presumptive histone tail-binding cleft. Mutation of the two residues Q908 and E909 abolished binding and may contribute to polar or ionic interactions with the positively charged ε-amino group of K4.

The conserved residues, aspartate, can be critical determinants for the CW domain's preference for N-terminal H3 tails, as in the ZCWPW1 CW, the N-terminal histone tail is interacting with the carbonyl oxygen of aspartate and in the ASHH2 CW an aspartate is placed in the same position. Each subfamily of the CW domains has a unique C-terminal extension with non-conserved residues. For example, in ZCWPW1 a third tryptophan present in the aromatic cage, while ASHH2 CW contains an amphipathic helix and when the ASHH2 CW domain was C-terminally truncated from residues M910, binding was lost. This can be interpreted as different CW domains showing different preferences for the three states of H3K4 methylation and the family specific C-terminal serve as determinants for recognition of the differently methylated H3K4 tails.

5. Concluding Remarks

In this study different biochemical, epigenetic and structural analyses approaches were used to identify the novel functions of SET domain proteins and their co-domains of *Arabidopsis thaliana.* Investigation of the biological function identified that SUVR4 HKMTase is involved in the repression of transcription for transposons (Paper I), while the ASHH2 protein required for normal plant development, is needed for maintenance of transcription levels of tissue-specific genes, and has a preference for actively transcribed genes with H3K4 methylation marks (Paper III). The further investigations of the co-domains of the SUVR4 and ASHH2 proteins showed that the co-domains are important for the function of these proteins. I have investigated the WIYLD domain of SUVR proteins, which has shown that the WIYLD domain specifically binds ubiquitin and histone H2Bub1, demonstrating a close connection between ubiquitin binding and H3K9 trimethylation. When ubiquitin binds the WIYLD domain, ubiquitin is stimulating the enzyme activity of SUVR4 and converts it from a strict dimethylase to a di/trimethylase *in vitro* (Paper I). The solution structure of the WIYLD domain revealed a four-helix bundle structure and confirms the binding of ubiquitin (Paper II). Amino acids responsible for the binding of S4WIYLD with ubiquitin were identified and the binding surface of S4WIYLD and ubiquitin was found from a docking analysis. SUVR4 was identified as the first HKMTase conferring H3K9me3, and in Arabidopsis H2Bub1 is associated with H3K9me3 in euchromatin. SUVR4 is only active on transposons, but based on the involvement of WIYLD Arg37 with the C-terminus of ubiquitin and H2Bub1 binding, it is tempting to hypothesize that other SUVR protein may be both readers and writers of the histone code by recognizing H2Bub1 and adding H3K9me3. ASHH2 is an H3K36me2/me3 HKMTase with preference for H3K4 methylation marks and its CW domain is binding mono- and dimethylated H3K4 peptides. Thus, ASHH2 is both reader and writer of histone code (Paper III). The solution structure of the ASHH2 CW domain contain a structural core built around two β-strands, a Zn^{2+} - binding sites and three short helical elements. Comparison between ZCWPW1 CW and ASHH2 CW shows that these domains show a similar histone tail binding site. As different CW domain recognize different states of H3K4 methylation, the structure suggest that the family specific C-terminal of the CW domain serve as determinant for the recognition of differently methylated H3K4 tails. Different epigenetic methods were used to identify the targets of the SUVR4 and ASHH2 proteins. It was found that the function of SUVR4 protein was dependent on the

cross-talk between different post-translational modifications of histone tails and the function of ASHH2 is depending on the CW domain, which acts as a reader of histone marks.

6. Future Perspectives

Current models suggest that the WIYLD domain is binding both free and conjugated ubiquitin, stimulating the activity of SUVR4. Mutational analyses and a docking model for S4WIYLD interacting with ubiquitin suggested that SUVR proteins might be acting both as readers and writers of the histone code. As no reader of H2Bub1 has been identified to date it will be exciting to investigate whether SUVR proteins can interact with H2Bub1 *in* vivo. Overexpression of SUVR1 and SUVR2 leads to smaller plants with an early flowering time compared to a wt plant (Silje V. Veiseth, unpublished data), but it remains to be shown whether any SUVR protein is involved in regulation of euchromatic genes. As the SUVR1, SUVR2 and SUVR4 protein are very similar, double or triple *suvr* mutants may be needed to test our hypothesis. NMR or crystallography technique can be used to solve the complex structure of WIYLD-ubiquitin, which will give more information about the mode of interaction of this complex. The mutation of more residues of the WIYLD domain, the binding surface mapping of ubiquitin by the titration of WIYLD domain with labeled ubiquitin and the mutation of the important residues of ubiquitin will give more information about the WIYLD-Ubiquitin interaction. A chromatin pull-down should be investigated to see if the WIYLD domain is also bind H2Bub in nucleosomal context. Furthermore, the full length SUVR4 structure with ubiquitin/histone tail peptides will provide knowledge about whether and how binding of ubiquitin might change the conformation of the SET domain to increase the enzymatic activity and to change the substrate and product specificity. I have shown that the WIYLD of SUVR1 and SUVR2 also binds ubiquitin, but I have not been able to demonstrate HKMTase activity for these proteins. A question to ask is what separates SUVR1 and SUVR2 from SUVR4, and one answer is the MET domain, a newly identified co-domain only present in SUVR1 and SUVR2, that needs to be investigated. For the CW project, an important question is whether the CW domain is sufficient for the targeting and necessary for the effector activity of CW proteins, e.g. the histone H3K36me3 methyltransferase activity of ASHH2, the putative binding to the methylated DNA of AtMBDs. It is also a challenge to elucidate the molecular basis for the different specificities towards mono-, di and tri-methylated H3K4. It would also be interesting to check if the plant and animal CW domains are interchangeable and can be used to change the targeting of given nuclear proteins and chromatin modifiers.

7. References

Alvarez-Venegas, R., and Avramova, Z. (2005). Methylation patterns of histone H3 Lys 4, Lys 9 and Lys 27 in transcriptionally active and inactive *Arabidopsis* genes and in *atx1* mutants. Nucleic acids research **33,** 5199-5207.

Alvarez-Venegas, R., Pien, S., Sadder, M., Witmer, X., Grossniklaus, U., and Avramova, Z. (2003). ATX-1, an *Arabidopsis* homolog of trithorax, activates flower homeotic genes. Current biology : CB **13,** 627-637.

Alvarez-Venegas, R., Sadder, M., Hlavacka, A., Baluska, F., Xia, Y., Lu, G., Firsov, A., Sarath, G., Moriyama, H., Dubrovsky, J.G., and Avramova, Z. (2006). The *Arabidopsis* homolog of trithorax, ATX1, binds phosphatidylinositol 5-phosphate, and the two regulate a common set of target genes. Proceedings of the national academy of sciences of the United States of America **103,** 6049-6054.

An, S., Yeo, K.J., Jeon, Y.H., and Song, J.J. (2011). Crystal structure of the human histone methyltransferase ASH1L catalytic domain and its implications for the regulatory mechanism. The journal of biological chemistry **286,** 8369-8374.

Ay, N., Irmler, K., Fischer, A., Uhlemann, R., Reuter, G., and Humbeck, K. (2009). Epigenetic programming via histone methylation at *WRKY53* controls leaf senescence in *Arabidopsis thaliana*. The plant journal : for cell and molecular biology **58,** 333-346.

Balciunas, D., and Ronne, H. (2000). Evidence of domain swapping within the jumonji family of transcription factors. Trends in biochemical sciences **25,** 274-276.

Bannister, A.J., and Kouzarides, T. (2005). Reversing histone methylation. Nature **436,** 1103-1106.

Baumbusch, L.O., Thorstensen, T., Krauss, V., Fischer, A., Naumann, K., Assalkhou, R., Schulz, I., Reuter, G., and Aalen, R.B. (2001). The *Arabidopsis thaliana* genome contains at least 29 active genes encoding SET domain proteins that can be assigned to four evolutionarily conserved classes. Nucleic acids research **29,** 4319-4333.

Berg, A., Meza, T.J., Mahic, M., Thorstensen, T., Kristiansen, K., and Aalen, R.B. (2003). Ten members of the *Arabidopsis* gene family encoding methyl-CpG-binding domain proteins are transcriptionally active and at least one, *AtMBD11*, is crucial for normal development. Nucleic acids research **31,** 5291-5304.

Berger, F., and Chaudhury, A. (2009). Parental memories shape seeds. Trends in plant science **14,** 550-556.

Berger, S.L. (2007). The complex language of chromatin regulation during transcription. Nature **447,** 407-412.

Bernatavichute, Y.V., Zhang, X., Cokus, S., Pellegrini, M., and Jacobsen, S.E. (2008). Genome-wide association of histone H3 lysine nine methylation with CHG DNA methylation in *Arabidopsis thaliana*. PloS one **3,** e3156.

Berr, A., Xu, L., Gao, J., Cognat, V., Steinmetz, A., Dong, A., and Shen, W.H. (2009). *SET DOMAIN GROUP25* encodes a histone methyltransferase and is involved in *FLOWERING LOCUS C* activation and repression of flowering. Plant physiology **151,** 1476-1485.

Berr, A., McCallum, E.J., Menard, R., Meyer, D., Fuchs, J., Dong, A., and Shen, W.H. (2010). *Arabidopsis SET DOMAIN GROUP2* is required for H3K4 trimethylation and is crucial for both sporophyte and gametophyte development. The plant cell **22,** 3232-3248.

Cartagena, J.A., Matsunaga, S., Seki, M., Kurihara, D., Yokoyama, M., Shinozaki, K., Fujimoto, S., Azumi, Y., Uchiyama, S., and Fukui, K. (2008). The *Arabidopsis* SDG4 contributes to the regulation of pollen tube growth by methylation of histone H3 lysines 4 and 36 in mature pollen. Developmental biology **315,** 355-368.

Cazzonelli, C.I., Cuttriss, A.J., Cossetto, S.B., Pye, W., Crisp, P., Whelan, J., Finnegan, E.J., Turnbull, C., and Pogson, B.J. (2009). Regulation of carotenoid composition and shoot branching in *Arabidopsis* by a chromatin modifying histone methyltransferase, SDG8. The plant cell **21,** 39-53.

Cerutti, H., and Casas-Mollano, J.A. (2009). Histone H3 phosphorylation: universal code or lineage specific dialects? Epigenetics : official journal of the DNA methylation Society **4,** 71-75.

Chang, Y.G., Song, A.X., Gao, Y.G., Shi, Y.H., Lin, X.J., Cao, X.T., Lin, D.H., and Hu, H.Y. (2006). Solution structure of the ubiquitin-associated domain of human BMSC-UbP and its complex with ubiquitin. Protein science : a publication of the protein society **15,** 1248-1259.

Chanvivattana, Y., Bishopp, A., Schubert, D., Stock, C., Moon, Y.H., Sung, Z.R., and Goodrich, J. (2004). Interaction of Polycomb-group proteins controlling flowering in *Arabidopsis*. Development **131,** 5263-5276.

Chen, M., Lv, S., and Meng, Y. (2010). Epigenetic performers in plants. Development, growth & differentiation **52,** 555-566.

Demidov, D., Hesse, S., Tewes, A., Rutten, T., Fuchs, J., Ashtiyani, R.K., Lein, S., Fischer, A., Reuter, G., and Houben, A. (2009). Aurora1 phosphorylation activity on histone H3 and its cross-talk with other post-translational histone modifications in *Arabidopsis*. The plant journal : for cell and molecular biology **59,** 221-230.

Dikic, I., Wakatsuki, S., and Walters, K.J. (2009). Ubiquitin-binding domains - from structures to functions. Nature reviews. Molecular cell biology **10,** 659-671.

Ding, Y., Avramova, Z., and Fromm, M. (2011). Two distinct roles of ARABIDOPSIS HOMOLOG OF TRITHORAX1 (ATX1) at promoters and within transcribed regions of ATX1-regulated genes. The plant cell **23,** 350-363.

Dong, G., Ma, D.P., and Li, J. (2008). The histone methyltransferase SDG8 regulates shoot branching in *Arabidopsis*. Biochemical and biophysical research communications **373,** 659-664.

Drag, M., Mikolajczyk, J., Bekes, M., Reyes-turcu, F.E., Ellman, J.A., Wilkinson, K.D., and Salvesen, G.S. (2008). Positional-scanning fluorigenic substrate libraries reveal unexpected specificity determinants of DUBs (deubiquitinating enzymes). The biochemical journal **415,** 367-375.

Ebbs, M.L., and Bender, J. (2006). Locus-specific control of DNA methylation by the *Arabidopsis* SUVH5 histone methyltransferase. The plant cell **18,** 1166-1176.

Fischer, A., Hofmann, I., Naumann, K., and Reuter, G. (2006). Heterochromatin proteins and the control of heterochromatic gene silencing in *Arabidopsis*. Journal of plant physiology **163,** 358-368.

Foreman, K.W., Brown, M., Park, F., Emtage, S., Harriss, J., Das, C., Zhu, L., Crew, A., Arnold, L., Shaaban, S., and Tucker, P. (2011). Structural and functional profiling of the human histone methyltransferase SMYD3. PloS one **6,** e22290.

Garner, T.P., Strachan, J., Shedden, E.C., Long, J.E., Cavey, J.R., Shaw, B., Layfield, R., and Searle, M.S. (2011). Independent Interactions of Ubiquitin-Binding Domains in a Ubiquitin-Mediated Ternary Complex. Biochemistry 50(42):9076-87.

Grini, P.E., Thorstensen, T., Alm, V., Vizcay-Barrena, G., Windju, S.S., Jorstad, T.S., Wilson, Z.A., and Aalen, R.B. (2009). The ASH1 HOMOLOG 2 (ASHH2) histone H3 methyltransferase is required for ovule and anther development in *Arabidopsis*. PloS one **4,** e7817.

Guo, L., Yu, Y., Law, J.A., and Zhang, X. (2010). SET DOMAIN GROUP2 is the major histone H3 lysine [corrected] 4 trimethyltransferase in *Arabidopsis*. Proceedings of the National academy of sciences of the United States of America **107,** 18557-18562.

Hammond-Martel, I., Yu, H., and Affar, E.B. (2011). Roles of ubiquitin signaling in transcription regulation. Cellular signalling. Cell Signal. 2012 Feb;24(2):410-21

Harper, J.W., and Schulman, B.A. (2006). Structural complexity in ubiquitin recognition. Cell **124,** 1133-1136.

He, F., Umehara, T., Saito, K., Harada, T., Watanabe, S., Yabuki, T., Kigawa, T., Takahashi, M., Kuwasako, K., Tsuda, K., Matsuda, T., Aoki, M., Seki, E., Kobayashi, N., Guntert, P., Yokoyama, S., and Muto, Y. (2010). Structural insight into the zinc finger CW domain as a histone modification reader. Structure **18,** 1127-1139.

Hicke, L., Schubert, H.L., and Hill, C.P. (2005). Ubiquitin-binding domains. Nature reviews. Molecular cell biology **6,** 610-621.

Hoppmann, V., Thorstensen, T., Kristiansen, P.E., Veiseth, S.V., Rahman, M.A., Finne, K., Aalen, R.B., and Aasland, R. (2011). The CW domain, a new histone recognition module in chromatin proteins. The EMBO journal **30,** 1939-1952.

Horn, P.J., and Peterson, C.L. (2002). Molecular biology. Chromatin higher order folding--wrapping up transcription. Science **297,** 1824-1827.

Hurley, J.H., Lee, S., and Prag, G. (2006). Ubiquitin-binding domains. The biochemical journal **399,** 361-372.

Ikeda, F., and Dikic, I. (2008). Atypical ubiquitin chains: new molecular signals. 'Protein Modifications: Beyond the Usual Suspects' review series. EMBO reports **9,** 536-542.

Inoue, N., Hess, K.D., Moreadith, R.W., Richardson, L.L., Handel, M.A., Watson, M.L., and Zinn, A.R. (1999). New gene family defined by MORC, a nuclear protein required for mouse spermatogenesis. Human molecular genetics **8,** 1201-1207.

Jackson, J.P., Lindroth, A.M., Cao, X., and Jacobsen, S.E. (2002). Control of CpNpG DNA methylation by the KRYPTONITE histone H3 methyltransferase. Nature **416,** 556-560.

Jackson, J.P., Johnson, L., Jasencakova, Z., Zhang, X., PerezBurgos, L., Singh, P.B., Cheng, X., Schubert, I., Jenuwein, T., and Jacobsen, S.E. (2004). Dimethylation of histone H3 lysine 9 is a critical mark for DNA methylation and gene silencing in *Arabidopsis thaliana*. Chromosoma **112,** 308-315.

Jacob, Y., Feng, S., LeBlanc, C.A., Bernatavichute, Y.V., Stroud, H., Cokus, S., Johnson, L.M., Pellegrini, M., Jacobsen, S.E., and Michaels, S.D. (2009). ATXR5 and ATXR6 are H3K27 monomethyltransferases required for chromatin structure and gene silencing. Nature structural & molecular biology **16,** 763-768.

Jacob, Y., Stroud, H., Leblanc, C., Feng, S., Zhuo, L., Caro, E., Hassel, C., Gutierrez, C., Michaels, S.D., and Jacobsen, S.E. (2010). Regulation of heterochromatic DNA replication by histone H3 lysine 27 methyltransferases. Nature **466,** 987-991.

Jenuwein, T., and Allis, C.D. (2001). Translating the histone code. Science **293,** 1074-1080.

Jenuwein, T., Laible, G., Dorn, R., and Reuter, G. (1998). SET domain proteins modulate chromatin domains in eu- and heterochromatin. Cellular and molecular life sciences : CMLS **54,** 80-93.

Jiang, D., Gu, X., and He, Y. (2009). Establishment of the winter-annual growth habit via *FRIGIDA*-mediated histone methylation at *FLOWERING LOCUS C* in *Arabidopsis*. The plant cell **21,** 1733-1746.

Jiang, D., Kong, N.C., Gu, X., Li, Z., and He, Y. (2011). *Arabidopsis* COMPASS-like complexes mediate histone H3 lysine-4 trimethylation to control floral transition and plant development. PLoS genetics **7,** e1001330.

Johnson, L.M., Law, J.A., Khattar, A., Henderson, I.R., and Jacobsen, S.E. (2008). SRA-domain proteins required for DRM2-mediated de novo DNA methylation. PLoS genetics **4,** e1000280.

Johnson, L.M., Bostick, M., Zhang, X., Kraft, E., Henderson, I., Callis, J., and Jacobsen, S.E. (2007). The SRA methyl-cytosine-binding domain links DNA and histone methylation. Current biology : CB **17,** 379-384.

Jones, B., Su, H., Bhat, A., Lei, H., Bajko, J., Hevi, S., Baltus, G.A., Kadam, S., Zhai, H., Valdez, R., Gonzalo, S., Zhang, Y., Li, E., and Chen, T. (2008). The histone H3K79 methyltransferase Dot1L is essential for mammalian development and heterochromatin structure. PLoS genetics **4,** e1000190.

Karachentsev, D., Sarma, K., Reinberg, D., and Steward, R. (2005). PR-Set7-dependent methylation of histone H4 Lys 20 functions in repression of gene expression and is essential for mitosis. Genes & development **19,** 431-435.

Karytinos, A., Forneris, F., Profumo, A., Ciossani, G., Battaglioli, E., Binda, C., and Mattevi, A. (2009). A novel mammalian flavin-dependent histone demethylase. The Journal of biological chemistry **284,** 17775-17782.

Khorasanizadeh, S. (2004). The nucleosome: from genomic organization to genomic regulation. Cell **116,** 259-272.

Ko, J.H., Mitina, I., Tamada, Y., Hyun, Y., Choi, Y., Amasino, R.M., Noh, B., and Noh, Y.S. (2010). Growth habit determination by the balance of histone methylation activities in Arabidopsis. The EMBO journal **29,** 3208-3215.

Kofler, M.M., and Freund, C. (2006). The GYF domain. The FEBS journal **273,** 245-256.

Komander, D. (2009). The emerging complexity of protein ubiquitination. Biochemical society transactions **37,** 937-953.

Komander, D. (2010). Mechanism, Specificity and Structure of the Deubiquitinases Conjugation and Deconjugation of Ubiquitin Family Modifiers, M. Groettrup, ed (Springer New York), pp. 69-87.

Komander, D., Clague, M.J., and Urbe, S. (2009). Breaking the chains: structure and function of the deubiquitinases. Nature reviews. Molecular cell biology **10,** 550-563.

Kouzarides, T. (2007). Chromatin modifications and their function. Cell **128,** 693-705.

Krichevsky, A., Kozlovsky, S.V., Gutgarts, H., and Citovsky, V. (2007a). *Arabidopsis* co-repressor complexes containing polyamine oxidase-like proteins and plant-specific histone methyltransferases. Plant signaling & behavior **2,** 174-177.

Krichevsky, A., Gutgarts, H., Kozlovsky, S.V., Tzfira, T., Sutton, A., Sternglanz, R., Mandel, G., and Citovsky, V. (2007b). C2H2 zinc finger-SET histone methyltransferase is a plant-specific chromatin modifier. Developmental biology **303,** 259-269.

Kwon, T., Chang, J.H., Kwak, E., Lee, C.W., Joachimiak, A., Kim, Y.C., Lee, J., and Cho, Y. (2003). Mechanism of histone lysine methyl transfer revealed by the structure of SET7/9-AdoMet. The EMBO journal **22,** 292-303.

Latham, J.A., and Dent, S.Y. (2007). Cross-regulation of histone modifications. Nature structural & molecular biology **14,** 1017-1024.

Lee, J.S., Smith, E., and Shilatifard, A. (2010). The language of histone crosstalk. Cell **142,** 682-685.

Liggins, A.P., Cooper, C.D., Lawrie, C.H., Brown, P.J., Collins, G.P., Hatton, C.S., Pulford, K., and Banham, A.H. (2007). MORC4, a novel member of the MORC family, is highly expressed in a subset of diffuse large B-cell lymphomas. British journal of haematology **138,** 479-486.

Liu, C., Lu, F., Cui, X., and Cao, X. (2010). Histone methylation in higher plants. Annual review of plant biology **61,** 395-420.

Luger, K., Mader, A.W., Richmond, R.K., Sargent, D.F., and Richmond, T.J. (1997). Crystal structure of the nucleosome core particle at 2.8 A resolution. Nature **389,** 251-260.

Makarevich, G., Leroy, O., Akinci, U., Schubert, D., Clarenz, O., Goodrich, J., Grossniklaus, U., and Kohler, C. (2006). Different Polycomb group complexes regulate common target genes in Arabidopsis. EMBO reports **7,** 947-952.

Marks, P., Rifkind, R.A., Richon, V.M., Breslow, R., Miller, T., and Kelly, W.K. (2001). Histone deacetylases and cancer: causes and therapies. Nature reviews. Cancer **1,** 194-202.

Mihola, O., Trachtulec, Z., Vlcek, C., Schimenti, J.C., and Forejt, J. (2009). A mouse speciation gene encodes a meiotic histone H3 methyltransferase. Science **323,** 373-375.

Morishita, M., and di Luccio, E. (2011). Structural insights into the regulation and the recognition of histone marks by the SET domain of NSD1. Biochemical and biophysical research communications **412,** 214-219.

Mueller, T.D., and Feigon, J. (2002). Solution structures of UBA domains reveal a conserved hydrophobic surface for protein-protein interactions. Journal of molecular biology **319,** 1243-1255.

Nathan, D., Ingvarsdottir, K., Sterner, D.E., Bylebyl, G.R., Dokmanovic, M., Dorsey, J.A., Whelan, K.A., Krsmanovic, M., Lane, W.S., Meluh, P.B., Johnson, E.S., and Berger, S.L. (2006). Histone sumoylation is a negative regulator in *Saccharomyces cerevisiae* and shows dynamic interplay with positive-acting histone modifications. Genes & development **20,** 966-976.

Naumann, K., Fischer, A., Hofmann, I., Krauss, V., Phalke, S., Irmler, K., Hause, G., Aurich, A.C., Dorn, R., Jenuwein, T., and Reuter, G. (2005). Pivotal role of AtSUVH2 in

heterochromatic histone methylation and gene silencing in *Arabidopsis*. The EMBO journal **24,** 1418-1429.

Ng, D.W., Wang, T., Chandrasekharan, M.B., Aramayo, R., Kertbundit, S., and Hall, T.C. (2007). Plant SET domain-containing proteins: structure, function and regulation. Biochimica et biophysica acta **1769,** 316-329.

Nickel, B.E., and Davie, J.R. (1989). Structure of polyubiquitinated histone H2A. Biochemistry **28,** 964-968.

Oh, S., Park, S., and van Nocker, S. (2008). Genic and global functions for Paf1C in chromatin modification and gene expression in Arabidopsis. PLoS genetics **4,** e1000077.

Ohno, A., Jee, J., Fujiwara, K., Tenno, T., Goda, N., Tochio, H., Kobayashi, H., Hiroaki, H., and Shirakawa, M. (2005). Structure of the UBA domain of Dsk2p in complex with ubiquitin molecular determinants for ubiquitin recognition. Structure **13,** 521-532.

Pena, P.V., Davrazou, F., Shi, X., Walter, K.L., Verkhusha, V.V., Gozani, O., Zhao, R., and Kutateladze, T.G. (2006). Molecular mechanism of histone H3K4me3 recognition by plant homeodomain of ING2. Nature **442,** 100-103.

Penengo, L., Mapelli, M., Murachelli, A.G., Confalonieri, S., Magri, L., Musacchio, A., Di Fiore, P.P., Polo, S., and Schneider, T.R. (2006). Crystal structure of the ubiquitin binding domains of rabex-5 reveals two modes of interaction with ubiquitin. Cell **124,** 1183-1195.

Perry, J., and Zhao, Y. (2003). The CW domain, a structural module shared amongst vertebrates, vertebrate-infecting parasites and higher plants. Trends in biochemical sciences **28,** 576-580.

Peterson, C.L., and Laniel, M.A. (2004). Histones and histone modifications. Current biology : CB **14,** R546-551.

Pien, S., Fleury, D., Mylne, J.S., Crevillen, P., Inze, D., Avramova, Z., Dean, C., and Grossniklaus, U. (2008). ARABIDOPSIS TRITHORAX1 dynamically regulates *FLOWERING LOCUS C* activation via histone 3 lysine 4 trimethylation. The plant cell **20,** 580-588.

Prasad, R., Zhadanov, A.B., Sedkov, Y., Bullrich, F., Druck, T., Rallapalli, R., Yano, T., Alder, H., Croce, C.M., Huebner, K., Mazo, A., and Canaani, E. (1997). Structure and expression pattern of human *ALR*, a novel gene with strong homology to ALL-1 involved in acute leukemia and to *Drosophila trithorax*. Oncogene **15,** 549-560.

Qian, C., and Zhou, M.M. (2006). SET domain protein lysine methyltransferases: Structure, specificity and catalysis. Cellular and molecular life sciences : CMLS **63,** 2755-2763.

Quina, A.S., Buschbeck, M., and Di Croce, L. (2006). Chromatin structure and epigenetics. Biochemical pharmacology **72,** 1563-1569.

Rajakumara, E., Law, J.A., Simanshu, D.K., Voigt, P., Johnson, L.M., Reinberg, D., Patel, D.J., and Jacobsen, S.E. (2011). A dual flip-out mechanism for 5mC recognition by the Arabidopsis SUVH5 SRA domain and its impact on DNA methylation and H3K9 dimethylation in vivo. Genes & development **25,** 137-152.

Rea, S., Eisenhaber, F., O'Carroll, D., Strahl, B.D., Sun, Z.W., Schmid, M., Opravil, S., Mechtler, K., Ponting, C.P., Allis, C.D., and Jenuwein, T. (2000). Regulation of chromatin structure by site-specific histone H3 methyltransferases. Nature **406,** 593-599.

Reyes-Turcu, F.E., Horton, J.R., Mullally, J.E., Heroux, A., Cheng, X., and Wilkinson, K.D. (2006). The ubiquitin binding domain ZnF UBP recognizes the C-terminal diglycine motif of unanchored ubiquitin. Cell **124,** 1197-1208.

Robinson, P.J., Fairall, L., Huynh, V.A., and Rhodes, D. (2006). EM measurements define the dimensions of the "30-nm" chromatin fiber: evidence for a compact, interdigitated structure. Proceedings of the National academy of sciences of the United States of America **103,** 6506-6511.

Roudier, F., Teixeira, F.K., and Colot, V. (2009). Chromatin indexing in *Arabidopsis*: an epigenomic tale of tails and more. Trends in genetics **25,** 511-517.

Roudier, F., Ahmed, I., Berard, C., Sarazin, A., Mary-Huard, T., Cortijo, S., Bouyer, D., Caillieux, E., Duvernois-Berthet, E., Al-Shikhley, L., Giraut, L., Despres, B., Drevensek, S., Barneche, F., Derozier, S., Brunaud, V., Aubourg, S., Schnittger, A., Bowler, C., Martin-Magniette, M.L., Robin, S., Caboche, M., and Colot, V. (2011). Integrative

epigenomic mapping defines four main chromatin states in *Arabidopsis*. The EMBO journal **30,** 1928-1938.

Ruthenburg, A.J., Li, H., Patel, D.J., and Allis, C.D. (2007). Multivalent engagement of chromatin modifications by linked binding modules. Nature reviews. Molecular cell biology **8,** 983-994.

Saleh, A., Alvarez-Venegas, R., and Avramova, Z. (2008). Dynamic and stable histone H3 methylation patterns at the Arabidopsis *FLC* and *AP1* loci. Gene **423,** 43-47.

Saleh, A., Al-Abdallat, A., Ndamukong, I., Alvarez-Venegas, R., and Avramova, Z. (2007). The Arabidopsis homologs of trithorax (ATX1) and enhancer of zeste (CLF) establish 'bivalent chromatin marks' at the silent *AGAMOUS* locus. Nucleic acids research **35,** 6290-6296.

Schapira, M. (2011). Structural Chemistry of Human SET Domain Protein Methyltransferases. Current chemical genomics **5,** 85-94.

Schatlowski, N., Stahl, Y., Hohenstatt, M.L., Goodrich, J., and Schubert, D. (2010). The CURLY LEAF interacting protein BLISTER controls expression of polycomb-group target genes and cellular differentiation of *Arabidopsis thaliana*. The plant cell **22,** 2291-2305.

Schotta, G., Ebert, A., Krauss, V., Fischer, A., Hoffmann, J., Rea, S., Jenuwein, T., Dorn, R., and Reuter, G. (2002). Central role of Drosophila SU(VAR)3-9 in histone H3-K9 methylation and heterochromatic gene silencing. The EMBO journal **21,** 1121-1131.

Schubert, D., Primavesi, L., Bishopp, A., Roberts, G., Doonan, J., Jenuwein, T., and Goodrich, J. (2006). Silencing by plant Polycomb-group genes requires dispersed trimethylation of histone H3 at lysine 27. The EMBO journal **25,** 4638-4649.

Seet, B.T., Dikic, I., Zhou, M.M., and Pawson, T. (2006). Reading protein modifications with interaction domains. Nature reviews. Molecular cell biology **7,** 473-483.

Shi, X., Hong, T., Walter, K.L., Ewalt, M., Michishita, E., Hung, T., Carney, D., Pena, P., Lan, F., Kaadige, M.R., Lacoste, N., Cayrou, C., Davrazou, F., Saha, A., Cairns, B.R., Ayer, D.E., Kutateladze, T.G., Shi, Y., Cote, J., Chua, K.F., and Gozani, O. (2006). ING2 PHD domain links histone H3 lysine 4 methylation to active gene repression. Nature **442,** 96-99.

Shiio, Y., and Eisenman, R.N. (2003). Histone sumoylation is associated with transcriptional repression. Proceedings of the national academy of sciences of the United States of America **100,** 13225-13230.

Sims, R.J., 3rd, Nishioka, K., and Reinberg, D. (2003). Histone lysine methylation: a signature for chromatin function. Trends in genetics : TIG **19,** 629-639.

Smith, R.D., and Walker, J.C. (1996). Plant protein phosphatases. Annual review of plant physiology and plant molecular biology **47,** 101-125.

Southall, S.M., Wong, P.S., Odho, Z., Roe, S.M., and Wilson, J.R. (2009). Structural basis for the requirement of additional factors for MLL1 SET domain activity and recognition of epigenetic marks. Molecular cell **33,** 181-191.

Spillane, C., MacDougall, C., Stock, C., Kohler, C., Vielle-Calzada, J.P., Nunes, S.M., Grossniklaus, U., and Goodrich, J. (2000). Interaction of the *Arabidopsis* polycomb group proteins FIE and MEA mediates their common phenotypes. Current biology : CB **10,** 1535-1538.

Sridhar, V.V., Kapoor, A., Zhang, K., Zhu, J., Zhou, T., Hasegawa, P.M., Bressan, R.A., and Zhu, J.K. (2007). Control of DNA methylation and heterochromatic silencing by histone H2B deubiquitination. Nature **447,** 735-738.

Strahl, B.D., and Allis, C.D. (2000). The language of covalent histone modifications. Nature **403,** 41-45.

Suzuki, M., Wang, H.H., and McCarty, D.R. (2007). Repression of the *LEAFY COTYLEDON 1/B3* regulatory network in plant embryo development by *VP1/ABSCISIC ACID INSENSITIVE 3-LIKE B3* genes. Plant physiology **143,** 902-911.

Tamada, Y., Yun, J.Y., Woo, S.C., and Amasino, R.M. (2009). *ARABIDOPSIS TRITHORAX-RELATED7* is required for methylation of lysine 4 of histone H3 and for transcriptional activation of *FLOWERING LOCUS C*. The plant cell **21,** 3257-3269.

Taverna, S.D., Li, H., Ruthenburg, A.J., Allis, C.D., and Patel, D.J. (2007). How chromatin-binding modules interpret histone modifications: lessons from professional pocket pickers. Nature structural & molecular biology **14,** 1025-1040.

Thorne, A.W., Sautiere, P., Briand, G., and Crane-Robinson, C. (1987). The structure of ubiquitinated histone H2B. The EMBO journal **6,** 1005-1010.

Thorstensen, T., Grini, P.E., and Aalen, R.B. (2011). SET domain proteins in plant development. Biochimica et biophysica acta **1809,** 407-420.

Thorstensen, T., Fischer, A., Sandvik, S.V., Johnsen, S.S., Grini, P.E., Reuter, G., and Aalen, R.B. (2006). The Arabidopsis SUVR4 protein is a nucleolar histone methyltransferase with preference for monomethylated H3K9. Nucleic acids research **34,** 5461-5470.

Thorstensen, T., Grini, P.E., Mercy, I.S., Alm, V., Erdal, S., Aasland, R., and Aalen, R.B. (2008). The Arabidopsis SET-domain protein ASHR3 is involved in stamen development and interacts with the bHLH transcription factor ABORTED MICROSPORES (AMS). Plant molecular biology **66,** 47-59.

Todi, S.V., Winborn, B.J., Scaglione, K.M., Blount, J.R., Travis, S.M., and Paulson, H.L. (2009). Ubiquitination directly enhances activity of the deubiquitinating enzyme ataxin-3. The EMBO journal **28,** 372-382.

Veiseth, S.V., Rahman, M.A., Yap, K.L., Fischer, A., Egge-Jacobsen, W., Reuter, G., Zhou, M.M., Aalen, R.B., and Thorstensen, T. (2011). The SUVR4 histone lysine methyltransferase binds ubiquitin and converts H3K9me1 to H3K9me3 on transposon chromatin in Arabidopsis. PLoS genetics **7,** e1001325.

Wang, D., Tyson, M.D., Jackson, S.S., and Yadegari, R. (2006). Partially redundant functions of two SET-domain polycomb-group proteins in controlling initiation of seed development in Arabidopsis. Proceedings of the national academy of sciences of the United States of America **103,** 13244-13249.

Wang, Y., Reddy, B., Thompson, J., Wang, H., Noma, K., Yates, J.R., 3rd, and Jia, S. (2009). Regulation of Set9-mediated H4K20 methylation by a PWWP domain protein. Molecular cell **33,** 428-437.

West, M.H., and Bonner, W.M. (1980). Histone 2B can be modified by the attachment of ubiquitin. Nucleic acids research **8,** 4671-4680.

Wilson, J.R., Jing, C., Walker, P.A., Martin, S.R., Howell, S.A., Blackburn, G.M., Gamblin, S.J., and Xiao, B. (2002). Crystal structure and functional analysis of the histone methyltransferase SET7/9. Cell **111,** 105-115.

Winget, J.M., and Mayor, T. (2010). The Diversity of Ubiquitin Recognition: Hot Spots and Varied Specificity. Molecular cell **38,** 627-635.

Wu, H., Min, J., Lunin, V.V., Antoshenko, T., Dombrovski, L., Zeng, H., Allali-Hassani, A., Campagna-Slater, V., Vedadi, M., Arrowsmith, C.H., Plotnikov, A.N., and Schapira, M. (2010). Structural biology of human H3K9 methyltransferases. PloS one **5,** e8570.

Wysocka, J., Swigut, T., Xiao, H., Milne, T.A., Kwon, S.Y., Landry, J., Kauer, M., Tackett, A.J., Chait, B.T., Badenhorst, P., Wu, C., and Allis, C.D. (2006). A PHD finger of NURF couples histone H3 lysine 4 trimethylation with chromatin remodelling. Nature **442,** 86-90.

Xiao, B., Jing, C., Wilson, J.R., Walker, P.A., Vasisht, N., Kelly, G., Howell, S., Taylor, I.A., Blackburn, G.M., and Gamblin, S.J. (2003). Structure and catalytic mechanism of the human histone methyltransferase SET7/9. Nature **421,** 652-656.

Xu, F., Zhang, K., and Grunstein, M. (2005). Acetylation in histone H3 globular domain regulates gene expression in yeast. Cell **121,** 375-385.

Xu, L., Zhao, Z., Dong, A., Soubigou-Taconnat, L., Renou, J.P., Steinmetz, A., and Shen, W.H. (2008). Di- and tri- but not monomethylation on histone H3 lysine 36 marks active transcription of genes involved in flowering time regulation and other processes in *Arabidopsis thaliana*. Molecular and cellular biology **28,** 1348-1360.

Xu, L., Menard, R., Berr, A., Fuchs, J., Cognat, V., Meyer, D., and Shen, W.H. (2009). The E2 ubiquitin-conjugating enzymes, AtUBC1 and AtUBC2, play redundant roles and are involved in activation of *FLC* expression and repression of flowering in *Arabidopsis thaliana*. The Plant journal : for cell and molecular biology **57,** 279-288.

Xu, S., Zhong, C., Zhang, T., and Ding, J. (2011). Structure of human lysine methyltransferase Smyd2 reveals insights into the substrate divergence in Smyd proteins. Journal of molecular cell biology **3,** 293-300.

Yadegari, R., Kinoshita, T., Lotan, O., Cohen, G., Katz, A., Choi, Y., Nakashima, K., Harada, J.J., Goldberg, R.B., Fischer, R.L., and Ohad, N. (2000). Mutations in the *FIE* and *MEA* genes that encode interacting polycomb proteins cause parent-of-origin effects on seed development by distinct mechanisms. The plant cell **12,** 2367-2382.

Yang, Z., Jiang, J., Stewart, D.M., Qi, S., Yamane, K., Li, J., Zhang, Y., and Wong, J. (2010). AOF1 is a histone H3K4 demethylase possessing demethylase activity-independent repression function. Cell research **20,** 276-287.

Yap, K.L., and Zhou, M.M. (2006). Structure and function of protein modules in chromatin biology. Results and problems in cell differentiation **41,** 1-23.

Yuan, X., Simpson, P., McKeown, C., Kondo, H., Uchiyama, K., Wallis, R., Dreveny, I., Keetch, C., Zhang, X., Robinson, C., Freemont, P., and Matthews, S. (2004). Structure, dynamics and interactions of p47, a major adaptor of the AAA ATPase, p97. The EMBO journal **23,** 1463-1473.

Zhang, K., Sridhar, V.V., Zhu, J., Kapoor, A., and Zhu, J.K. (2007). Distinctive core histone post-translational modification patterns in *Arabidopsis thaliana*. PloS one **2,** e1210.

Zhang, X., Bernatavichute, Y.V., Cokus, S., Pellegrini, M., and Jacobsen, S.E. (2009). Genome-wide analysis of mono-, di- and trimethylation of histone H3 lysine 4 in *Arabidopsis thaliana*. Genome biology **10,** R62.

Zhang, X., Tamaru, H., Khan, S.I., Horton, J.R., Keefe, L.J., Selker, E.U., and Cheng, X. (2002). Structure of the Neurospora SET domain protein DIM-5, a histone H3 lysine methyltransferase. Cell **111,** 117-127.

Zhang, Y., and Reinberg, D. (2001). Transcription regulation by histone methylation: interplay between different covalent modifications of the core histone tails. Genes & development **15,** 2343-2360.

Zhao, Z., Yu, Y., Meyer, D., Wu, C., and Shen, W.H. (2005). Prevention of early flowering by expression of *FLOWERING LOCUS C* requires methylation of histone H3 K36. Nature cell biology **7,** 1256-1260.

Zhou, D.X. (2009). Regulatory mechanism of histone epigenetic modifications in plants. Epigenetics : official journal of the DNA methylation society **4,** 15-18.

I

OPEN ACCESS Freely available online PLoS GENETICS

The SUVR4 Histone Lysine Methyltransferase Binds Ubiquitin and Converts H3K9me1 to H3K9me3 on Transposon Chromatin in Arabidopsis

Silje V. Veiseth[1], Mohummad A. Rahman[1], Kyoko L. Yap[2], Andreas Fischer[3], Wolfgang Egge-Jacobsen[1,4], Gunter Reuter[3], Ming-Ming Zhou[2], Reidunn B. Aalen[1], Tage Thorstensen[1¤*]

1 Department of Molecular Biosciences, University of Oslo, Oslo, Norway, **2** Department of Structural and Chemical Biology, Mount Sinai School of Medicine, New York, New York, United States of America, **3** Institute of Biology, Developmental Genetics, Martin Luther University Halle, Halle, Germany, **4** Glyconor Mass Spectrometry, Department of Molecular Biosciences, University of Oslo, Oslo, Norway

Abstract

Chromatin structure and gene expression are regulated by posttranslational modifications (PTMs) on the N-terminal tails of histones. Mono-, di-, or trimethylation of lysine residues by histone lysine methyltransferases (HKMTases) can have activating or repressive functions depending on the position and context of the modified lysine. In Arabidopsis, trimethylation of lysine 9 on histone H3 (H3K9me3) is mainly associated with euchromatin and transcribed genes, although low levels of this mark are also detected at transposons and repeat sequences. Besides the evolutionarily conserved SET domain which is responsible for enzyme activity, most HKMTases also contain additional domains which enable them to respond to other PTMs or cellular signals. Here we show that the N-terminal WIYLD domain of the Arabidopsis SUVR4 HKMTase binds ubiquitin and that the SUVR4 product specificity shifts from di- to trimethylation in the presence of free ubiquitin, enabling conversion of H3K9me1 to H3K9me3 *in vitro*. Chromatin immunoprecipitation and immunocytological analysis showed that SUVR4 *in vivo* specifically converts H3K9me1 to H3K9me3 at transposons and pseudogenes and has a locus-specific repressive effect on the expression of such elements. Bisulfite sequencing indicates that this repression involves both DNA methylation–dependent and –independent mechanisms. Transcribed genes with high endogenous levels of H3K4me3, H3K9me3, and H2Bub1, but low H3K9me1, are generally unaffected by SUVR4 activity. Our results imply that SUVR4 is involved in the epigenetic defense mechanism by trimethylating H3K9 to suppress potentially harmful transposon activity.

Citation: Veiseth SV, Rahman MA, Yap KL, Fischer A, Egge-Jacobsen W, et al. (2011) The SUVR4 Histone Lysine Methyltransferase Binds Ubiquitin and Converts H3K9me1 to H3K9me3 on Transposon Chromatin in Arabidopsis. PLoS Genet 7(3): e1001325. doi:10.1371/journal.pgen.1001325

Editor: Li-Jia Qu, Peking University, China

Received August 18, 2010; **Accepted** February 3, 2011; **Published** March 10, 2011

Funding: This work was supported by the Research Council of Norway (http://www.forskningsradet.no/), projects no. 146652/431 and 183609/S10, and the Deutsche Forschungsgemeinschaft (http://www.dfg.de) (DFG, SFB 648). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: tageth@imbv.uio.no

¤ Current address: The Norwegian Biotechnology Advisory Board, Oslo, Norway

Introduction

In eukaryotes, gene expression and chromatin structure is specified by the combinatorial pattern of posttranslational modifications (PTMs) on the histone tails, which include phosphorylation, acetylation, methylation, SUMOylation and ubiquitination [1,2]. These PTMs are interdependent, thus providing regulatory cross-talk, and established at the histone tails in a coordinated manner by different classes of highly specific chromatin modifying enzymes.

The combination of PTMs constitutes the so-called histone code, and their downstream effect on chromatin organization and gene expression is mediated by nonhistone effector proteins that contain domains that bind or "read" this code in order to specify epigenetic function. Such domains show specificity for particular modified residues (e.g. acetylation or methylation of lysine) in the context of its surrounding amino acid sequence, and for the state of the modification (e.g. H3K9me1 vs H3K9me3) [1,3]. For example, domains belonging to the Royal Superfamily, including the chromodomain, Tudor domain and MBT domain and members of the PHD finger family, bind methylated lysine residues on the histone tails [4]. More specifically, the PHD finger of the ORC1 protein in Arabidopsis binds H3K4me3, but not H3K4me1 or H3K4me2 at target genes, and this mediates H4K20 trimethylation and activates transcription [5].

Lysine ubiquitination of histones and other target proteins is a three step process involving Ub (ubiquitin)-activating (E1), Ub-conjugating (E2) and Ub-ligating (E3) enzymes, eventually leading to monoubiquitination, multi-monoubiquitination or polyubiquitination [6,7]. Ubiquitin binding domains (UBDs) represent a new class of motifs that enable proteins to bind non-covalently to the PTM ubiquitin. More than twenty families have been identified to date, and they differ in structure and the type of ubiquitin modification they recognize [6,7]. Poly-Ub chains linked via the K48 residue of ubiquitin are largely recognized by UBDs of receptors that target proteins for proteosomal degradation, while monoubiquitin is recognized by UBDs of proteins involved in processes like DNA repair, regulation of protein activity, chromatin remodeling and transcription [6–8].

Author Summary

The characteristics of the diverse cell types in multicellular organisms result from differential gene expression that is dependent on the level of DNA packaging. Genes that are essential for the function of the cell are expressed; while unessential genes, and DNA elements (transposons or "jumping genes") that can move from one position to another within a genome and potentially cause deleterious mutations, are repressed. The mechanisms evolved in eukaryotes to avoid unwanted gene expression and transposon movement include DNA methylation and specific combinations of post translational modifications (PTMs) of the histones that package DNA. Here we show that the SUVR4 enzyme binds the signaling protein ubiquitin and that ubiquitin enables the enzyme to trimethylate lysine 9 (H3K9me3) of histone H3. In contrast to other reports demonstrating an activating role on expressed genes, we show that H3K9me3 has a locus-specific repressive effect on the expression of transposons. The specificity is maintained by the communication with other PTMs on transposons and euchromatic genes, which has a stimulating or repressing effect on enzyme activity, respectively. Our results demonstrate how repression of transcription can be restricted to specific targets and demonstrate that this repression involves a context-dependent read-out of different PTMs.

The cross-talk between H2B monoubiquitination (H2Bub1) and histone methylation has been extensively studied and is highly conserved from yeast to human. These studies show that monoubiquitination of H2B recruits proteins that direct histone H3K4 di- and trimethylation but not monomethylation by activation of the Set1 histone lysine methyltransferase (HKMTase) of the COMPASS complex (reviewed in [9,10]). In Arabidopsis, H2B monoubiquitination at K143 coincides with active transcription [11–13]. Deubiquitinating enzymes (DUBs) oppose the function of E3 ligases by deubiquitinating Ub-conjugated proteins. Increased H2Bub1 caused by a mutation in the DUB SUP32/UBP26, leads to reduced H3K9me2 and increased H3K4me3 at transposons that correlate with increased transcription [11]. A key function for DUBs is to generate a pool of free ubiquitin monomers from ubiquitin precursors synthesized from Ub-encoding genes, and from polyubiquitin chains and ubiquitin conjugates [14]. Free monomeric ubiquitin is required under stress conditions, and organisms defective in ubiquitin precursor proteins or DUBs are more sensitive to stress. In yeast, heat stress stimulates the production and activation of the Doa4 deubiquitinase which increases the supply of free monomeric ubiquitin by cleaving polyubiquitin [15].

HKMTases contain SET domains with specificities for different lysine residues on the histone tails, and may be involved in either gene activation or gene repression depending on which lysine residue is methylated [16]. In general, methylation of H3K9, H3K27 and H4K20 has been associated with heterochromatin and gene repression, while H3K4, H3K36 and H3K79 methylation has been related to euchromatin and gene activation [1]. The downstream effect of histone methylation also depends on the number of methyl groups at each lysine residue. Histones mono-, di-, or trimethylated at lysines are differently distributed within eu- and heterochromatin, each potentially indexing a specific biological outcome [17,18]. For example, in Arabidopsis, H3K36 trimethylation, but not H3K36 monomethylation, shows a strong positive correlation with transcription of MADS box genes involved in flowering-time and flower development [19,20].

Although lysine methylation to a large extent is conserved between eukaryotes, the distribution and biological outcome of the methylation may be different. H3K9me1, H3K9me2 and H3K27me2 are for instance predominantly found in the chromocenters of Arabidopsis but not in mouse chromocenters (reviewed in [21,22]). Conversely, H3K9me3 and H4K20me3 that localize to heterochromatin in mouse are mainly associated with euchromatin in Arabidopsis. Additionally, recent results suggest that in contrast to other eukaryotes, H3K9me3 methylation correlates with gene transcription and might have a slight activating function in Arabidopsis [23,24].

H3K9 methylation is carried out by proteins of the SU(VAR)3-9 subgroup which consists of 14 proteins in Arabidopsis; the SU(VAR) 3-9 HOMOLOGs SUVH1-SUVH9, and the more distantly related SU(VAR) 3-9 RELATED proteins SUVR1-5 [25]. In addition to the SET domain the SUVH proteins contain the YDG/SRA domain that has been shown to bind methylated DNA and might direct SUVH mediated H3K9me2 to heterochromatin or stimulate its activity [26]. Thus in Arabidopsis, the SUVH proteins link the epigenetic gene-silencing marks H3K9me2 and DNA-methylation and work as transcriptional repressors of transposons or inverted repeat sequences, for instance by directing CHG methylation via the CMT3 DNA methyltransferase (reviewed in [27]). In contrast to the SUVH proteins, the SUVR1, SUVR2 and SUVR4 proteins do not contain an YDG/SRA domain, but an N-terminal WIYLD domain of unknown function [28], suggesting another mode of action for these proteins. SUVR proteins associate with the nucleolus or euchromatin, and we have earlier shown that SUVR4 can dimethylate H3K9 when this position is monomethylated [28].

In the present study we show that the WIYLD domain of SUVR4 specifically binds ubiquitin, demonstrating a close connection between ubiquitin binding and histone H3K9 methylation. We have furthermore revealed that ubiquitin stimulates the enzyme activity of SUVR4 and converts SUVR4 from a strict dimethylase to a di/trimethylase *in vitro*. Chromatin Immunoprecipitation (ChIP) analysis of Arabidopsis lines with reduced or enhanced expression of *SUVR4*, demonstrate that SUVR4 localizes to both euchromatin and heterochromatin *in vivo*, but only converts H3K9me1 to H3K9me3 at transposons and pseudogenes. SUVR4 dependent H3K9 trimethylation correlates with locus specific transcriptional repression of transposable elements intercalated within euchromatin of the Arabidopsis genome.

Results

The WIYLD domain is a ubiquitin-binding domain

To address the function of the SUVR4 WIYLD domain, a construct encompassing only this domain (Figure 1A) was used in a yeast two-hybrid screen to identify interacting proteins. One positive clone identified in this screen, contained the full-length coding sequence (CDS) of UBIQUITIN EXTENSION PROTEIN 1 (UBQ1, AT3G52590) (Figure 1B). The UBQ1 protein consists of an N-terminal ubiquitin moiety and the C-terminal ribosomal protein L40 [29]. These moieties were subcloned and tested separately for their interaction with SUVR4-WIYLD. Clones containing the ubiquitin moiety, but not clones containing the L40 moiety, supported growth on selective media when transformed into yeast cells and mated with cells containing SUVR4-WIYLD, suggesting that SUVR4 specifically interacts with ubiquitin (Figure 1B). This was confirmed in an *in vitro* pull-down experiment, where SUVR4-WIYLD pulled down full-length UBQ1 and ubiquitin but not L40 (Figure 1C).

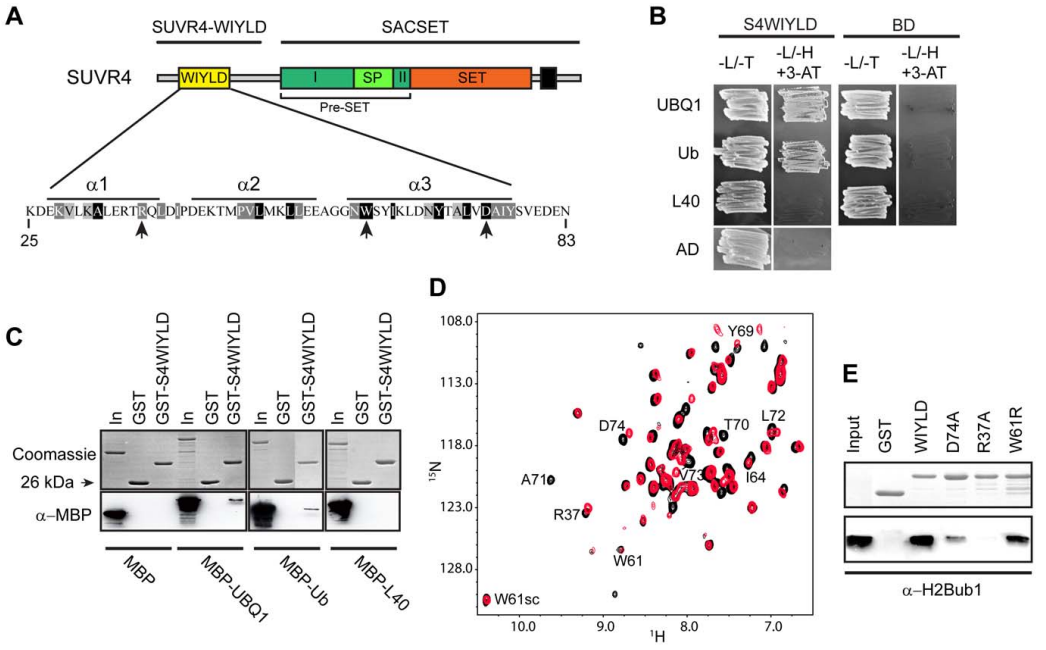

Figure 1. The WIYLD domain is a ubiquitin-binding domain. (A) Layout of the SUVR4 full-length protein, showing the different domains/motifs and the regions included in the two constructs SUVR4-WIYLD and SACSET. SP, SUVR pre-SET; I, pre-SET I; II, pre-SET II; SET, SET domain. Black box indicates the post-SET domain [28].The amino acid sequence of SUVR4 from N_{25} to K_{83} encompassing the WIYLD domain, with conserved residues shaded in black, and residues mutated in this work indicated with arrowheads. (B) Yeast two-hybrid interaction test between SUVR4-WYILD and the full-length CDS of UBQ1, as well as the N-terminal ubiquitin (Ub) and the C-terminal ribosomal L40 moieties of UBQ1. –L/-T – medium selective for diploid colonies; -L/-T/-H +3 AT – medium selective for protein-protein interactions. AD, control mating with empty prey vector; BD, control matings with empty bait vector. (C) GST-SUVR4-WIYLD immobilized on glutathione sepharose beads were used to pull down MBP, MBP-UBQ1, MBP-Ub or MBP-L40 from bacterial lysate. Pull-down reactions were separated on SDS-PAGE, blotted onto a PVDF membrane and detected with an anti-MBP antibody. IN, input (5%); GST, GST negative control. (D) [^{1}H,^{15}N]-HSQC spectrum of SUVR4-WIYLD in its free form (black), and after the addition of excess ubiquitin to a molar ratio of 1:3 (red). The assigned amino acid residues are indicated. (E) GST pull-down of H2Bub1. GST-SUVR4-WIYLD was mutated at positions D74, R37 and W61 and used for GST pull-down of core histones from calf thymus. The pull down reactions were blotted onto a PVDF membrane and probed with an antibody against ubiquitinylated H2B (H2Bub1).
doi:10.1371/journal.pgen.1001325.g001

To address whether the WIYLD domain binds ubiquitin in its unconjugated form and to identify residues directly involved in the interaction between WIYLD and ubiquitin, an NMR analysis was performed. The [^{1}H,^{15}N]-HSQC spectrum of ^{15}N-isotopically labeled SUVR4-WIYLD is well-dispersed demonstrating that the protein domain is folded (Figure 1D). Upon titration of ubiquitin, chemical shift perturbations were observed for a number of residues including the six consecutive amino acids Y_{69}TALVD$_{74}$ of helix 3 (Figure 1D), indicating that they are involved in binding. Alignment of SUVR4-WIYLD with WIYLD domains in other proteins have earlier shown that many of these residues are highly conserved (Figure 1A and [28]).

SUVR4 binds and efficiently methylates calf thymus histone H3 as well as H3K9me1 peptides *in vitro*, but shows only weak activity against recombinant histones, arguing that SUVR4 cross-talks to premodified histones [28]. Since the WIYLD domain binds ubiquitin, and SUVR4 binds and methylates histones, we tested whether the WIYLD domain binds H2B monoubiquitinated on lysine 143 (H2Bub1), which is the only ubiquitination on core histones reported so far in Arabidopsis [11,30]. In these experiments the WIYLD domain indeed was able to pull down H2Bub1, however, when R37 and D74 were mutated, the interaction was strongly reduced (Figure 1E). This supports the chemical shift perturbations shown by the NMR analysis, arguing that these residues are directly involved in ubiquitin binding. Interestingly, the invariant W61 residue that showed no shift in the NMR analysis, only weakly affected the WIYLD-ubiquitin interaction when mutated, confirming that this position is not crucial for ubiquitin binding.

The WIYLD domain enhances the HKMTase activity of SUVR4 through binding of ubiquitin

As the WIYLD domain was able to bind ubiquitin (Figure 1D), we asked whether ubiquitin could stimulate SUVR4 enzyme activity, as previously shown for the deubiquitinase USP5 [31]. To this end, we compared the activity of a SUVR4 protein without the WIYLD domain to a full-length SUVR4 protein, both in fusion with the Maltose Binding Protein (MBP-SACSET and MBP-SUVR4, Figure 1A), with and without the addition of ubiquitin. In both cases the full-length protein showed higher enzymatic activity than the truncated SACSET fragment (Figure 2A, B), suggesting that the WIYLD domain has a positive

effect on the catalytic activity of SUVR4 although the domain itself does not contain HMTase activity (Figure S1C). The difference in activity was more pronounced when ubiquitin was added to the reaction. With ubiquitin the full-length protein was stimulated 2-3 fold whereas the SACSET construct was only weakly affected, suggesting that most of the ubiquitin response is mediated through the WIYLD domain (Figure 2A, B). Addition of free ubiquitin only stimulates enzymatic activity of the SUVR4 protein on histone H3 but does not affect its specificity as no other core histones becomes methylated (Figure 2C).

Ubiquitin converts SUVR4 from a strict dimethylase to a di/trimethylase

Using H3K9me1 and H3K9me2 peptides we tested whether the increased SUVR4 enzyme activity after the addition of ubiquitin also affected the product specificity. As expected from previous results [28], H3K9me1 peptides were the preferred substrate as unmethylated peptides were only weakly methylated (Figure S1A), and no activity against H3K9me2 peptides was observed in the absence of ubiquitin. Methylation of H3K9me1 modified peptides was increased 2.5–3 fold when ubiquitin was added to the reaction (Figure 2D). Unexpectedly we also observed methylation of the H3K9me2 peptide in the presence of ubiquitin, suggesting that ubiquitin converted the SUVR4 protein to a histone H3K9 trimethylase (Figure 2D, Figure S1B). The activity on H3K9me2 peptides was however several folds lower than when H3K9me1 peptides were used. No activity was observed on H3K9me3 peptides either with or without ubiquitin, excluding the possibility that any other lysine of histone H3 1-21 was methylated by SUVR4, underscoring the specificity against H3K9 (Figure 2D).

The products from the enzyme reactions using peptide substrates were analyzed by peptide mass fingerprinting. After 3 hours incubation, the reactions containing SUVR4 only converted 40.9% of the H3K9me1 peptide to H3K9me2, while 0% was converted to H3K9me3 (Figure 2E, upper middle panel). In the reactions containing ubiquitin, 90.2% of the H3K9me1 peptide was converted to H3K9me2 while 3.5% was converted to H3K9me3 (Figure 2E, upper right panel). When H3K9me2 peptides were used as substrate, we did not see any conversion to H3K9me3 above background level in the absence of ubiquitin (3%

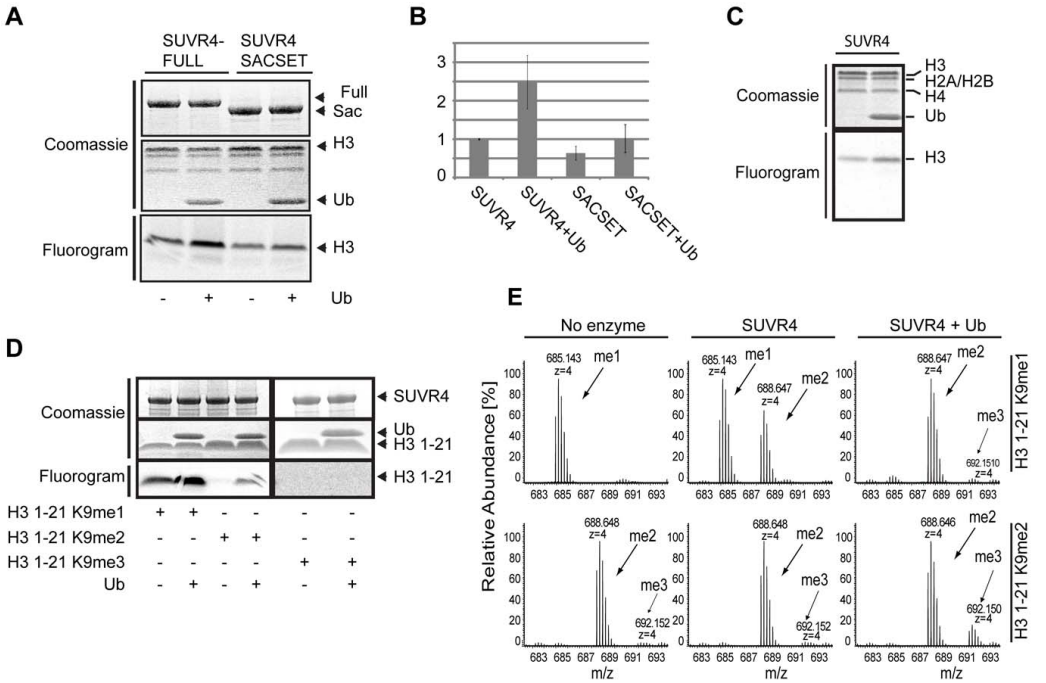

Figure 2. SUVR4 HKMTase activity is stimulated by free ubiquitin *in vitro*. (A) HKMTase assay on core histones using a construct encompassing the SACSET domain of SUVR4 or the full-length SUVR4 protein without and with the addition of free ubiquitin. (B) Quantification of band intensity from fluorogram in A, relative to the reaction with SUVR4 without adding ubiquitin. The graph represents the average of four independent assays. (C) HKMTase assay with SUVR4 full-length using core histones from calf thymus as substrate, without (left) and with (right) the addition of 5 μg free ubiquitin, respectively. (D) The same assay as (C) but using histone H3 1-21 K9me1, H3 1-21 K9me2 or H3 1-21 K9me3 peptides with and without the addition of 5 μg free ubiquitin. (E) Peptide mass fingerprints of the products of an identical HKMTase assay as in C, using unlabelled SAM as methyl donor and H3 1-21 K9me1 (upper panel) or H3 1-21 K9me2 peptides as substrate (lower panel). Products from assays without (left) the addition of SUVR4 enzyme, containing SUVR4 protein (middle) and SUVR4 protein with the addition of 5 μg ubiquitin (right), were analyzed. The mass spectra of each peptide are shown as bars representing the mass-to-charge ratio (m/z), and the most abundant m/z is set to 100%. The length of the bars indicates abundance of the m/z relative to the most abundant. All enzyme assays were repeated at least 4 times with independent protein samples.

doi:10.1371/journal.pgen.1001325.g002

background H3K9me3, versus 3.5% when SUVR4 was added to the reaction) (Figure 2E, lower middle panel), however when ubiquitin was present together with SUVR4, a 16.4% conversion from H3K9me2 to H3K9me3 was found (Figure 2E, lower right panel). This suggests that ubiquitin stimulates the catalytic activity of SUVR4 and alters the product specificity in that it converts SUVR4 from a strict dimethylase to a di/trimethylase.

SUVR4 directs H3K9 trimethylation to transposon chromatin

As SUVR4 converts H3K9me1 to H3K9me2/3 *in vitro*, we asked how these modifications were affected by SUVR4 *in vivo*. Since no SUVR4 T-DNA knock-out insertion lines were available, knock-down RNAi lines for SUVR4 were established. We also generated GFP overexpression (OE) lines where SUVR4-GFP expression was driven by the strong constitutive 35S promoter, giving a uniform SUVR4-distribution in the nucleus in addition to accumulation in the nucleolus or in foci of unknown function (Figure S2). A weaker glucocorticoid-inducible construct has earlier been reported to give an almost exclusive nucleolar localization of SUVR4 [28]. We did not observe any phenotypes under the tested growing conditions for neither the SUVR4-GFP line, nor the SUVR4 RNAi line.

H3K9me1-3 display different nuclear distributions, with high H3K9me1/2 in chromocenters and pericentric heterochromatin, whereas H3K9me3 is distributed more uniformly in the nucleoplasm with highest concentration in euchromatin and at expressed genes [32]. Immunocytological analysis on seedling leaves using specific antibodies against H3K9me showed a strong reduction in H3K9me1 and a corresponding increase in H3K9me3 in nuclei with high SUVR4-GFP expression (Figure 3A). Nuclei from lines with a low SUVR4-GFP expression did not show this effect on H3K9me1 and H3K9me3 methylation, suggesting that the global changes in H3K9me1 and H3K9me3 correlated with SUVR4-GFP expression (Figure 3A).

To analyze this effect at individual genes, ChIP experiments were performed with the same antibodies as used for immunocytological analysis and an antibody specific for GFP, respectively. Different classes of transposon sequences were selected for ChIP analysis, as these sequences are likely targets of SUVR4 because of their high H3K9me1 level (Figure 3B and Table 1). These experiments confirmed that SUVR4 is associated with transposons and genes both in eu- and heterochromatin, but a significantly higher amount of SUVR4-GFP is found at euchromatic genes like *TUB8* and *ACTIN2* (Figure S3). However, only transposon and pseudogenes like *AtSN1, AtGP1, AtMU1, AtCOPIA4* and *MULE At2g15810* were affected by overexpression of SUVR4, resulting in a drastic increase in H3K9me3 and reduction of H3K9me1 (Figure 3B). We did not see any effect of SUVR4 OE for highly expressed genes like *TUB8* or *ACTIN2*, or for the moderately expressed transposon At4g13120, all with an already low level of H3K9me1. Although having a dramatic effect on H3K9me3 at transposons, SUVR4 OE did not affect the distribution of the euchromatic mark H2Bub1 at any of the tested sequences (Figure S4A).

As the *35S* driven SUVR4-GFP construct could lead to unspecific downstream effects due to ectopic and elevated SUVR4 expression, we complemented the OE data with ChIP analysis of two of the transposons in knock-down SUVR4 RNAi plants. The RNAi lines showed a 90% reduction of the SUVR4 expression level compared to wild type (Figure S5 A). In contrast to the OE line, there was an increase of H3K9me1 on *AtSN1* and *MULE At2g15810* (Figure 3C). Furthermore, there was a corresponding reduction of H3K9me3, suggesting that SUVR4 directs H3K9me3 methylation on transposons. The weak reduction of H3K9me3 could reflect the residual SUVR4 expression in the RNAi line and possibly redundancy with other H3K9me3 methyltransferases at these sequences. Together, these data suggest that although SUVR4 is localized in both eu- and heterochromatin, it is active only on target sequences with a high level of H3K9me1, where its activity increases H3K9me3 at the expense of the H3K9me1 level.

H3K4me3 reduces SUVR4 HKMTase activity

Recent studies suggest that in Arabidopsis H3K9me3 associates with euchromatin and transcriptional activation of genes [23,24,32]. In contrast, H3K9me1 is a mark mainly associated with repetitive sequences in chromocenters and pericentric heterochromatin in Arabidopsis [21]. The specific activity of SUVR4 on transposon chromatin although associated with both transposons and euchromatic genes (Figure 3, S3), made us speculate that the lack of SUVR4 activity on euchromatic genes was due to cross-talk to PTMs characteristic for euchromatin. We thus tested histone tail peptides that were mono- or trimethylated at H3K4 but devoid of H3K9me in an *in vitro* HKMTase assay (Figure 4). SUVR4 activity was not affected by monomethyl H3K4, whereas trimethyl H3K4 reduced SUVR4 activity significantly (Figure 4 A, B), arguing that chromatin associated with genes like *TUB8* and *ACTIN2*, with a high level of this mark, might not be good substrate for SUVR4 activity.

SUVR4 is a transcriptional repressor of transposable elements

To evaluate the effect of SUVR4 mediated H3K9me3 methylation on transposon transcription we investigated the expression of three of the ChIP-analyzed transposons, *MULE At2g15810*, *AtIS112A* (At4g04293) and *AtCOPIA4*, which all had a high level of H3K9me1 and were expressed in wild type plants (Figure 3B, C, Figure S5 B and Table 1). In the OE line, all the studied transposons showed significant reduction in expression compared to wild type (60%, 80% and 35%, respectively, Figure 5A), suggesting that SUVR4 acts as a repressor of these transposable elements. As a control, we used the At4g13120 transposable element of intermediate expression with a very low H3K9me1 level which is not a target of SUVR4 methylation (Figure 3, Figure 5A and Table 1). This transposon was also unaffected in its transcription level in SUVR4-GFP overexpression lines.

In the RNAi line we did not see a corresponding release of repression for the *AtCOPIA4* and *AtIS112A* elements, however, the *MULE At2g15810* element was induced 2.5 to 3- fold in the RNAi line compared to wild type (Figure 5A). Interestingly, the gene *Cyp40* which is known to be regulated by MULE [33] showed the same expression response to SUVR4 as *MULE At2g15810*, although weaker (Figure 5A). The *AtSN1* repeat interspersed within euchromatin, and the heterochromatin localized *AtMU1* that are silent in wild type plants (Table 1 and Figure S5 B), were examined in both the RNAi and OE line but we did not detect any signal above the –RT control reaction, arguing that these transposons were not reactivated in any of the lines (data not shown).

SUVR4 shows a locus-specific effect on DNA methylation

H3K9me2 directed by SUVH proteins regulates non-CG methylation in Arabidopsis [34]. To determine if there was a similar correlation between DNA methylation and the H3K9me3 methylation directed by SUVR4, bisulfite sequencing was

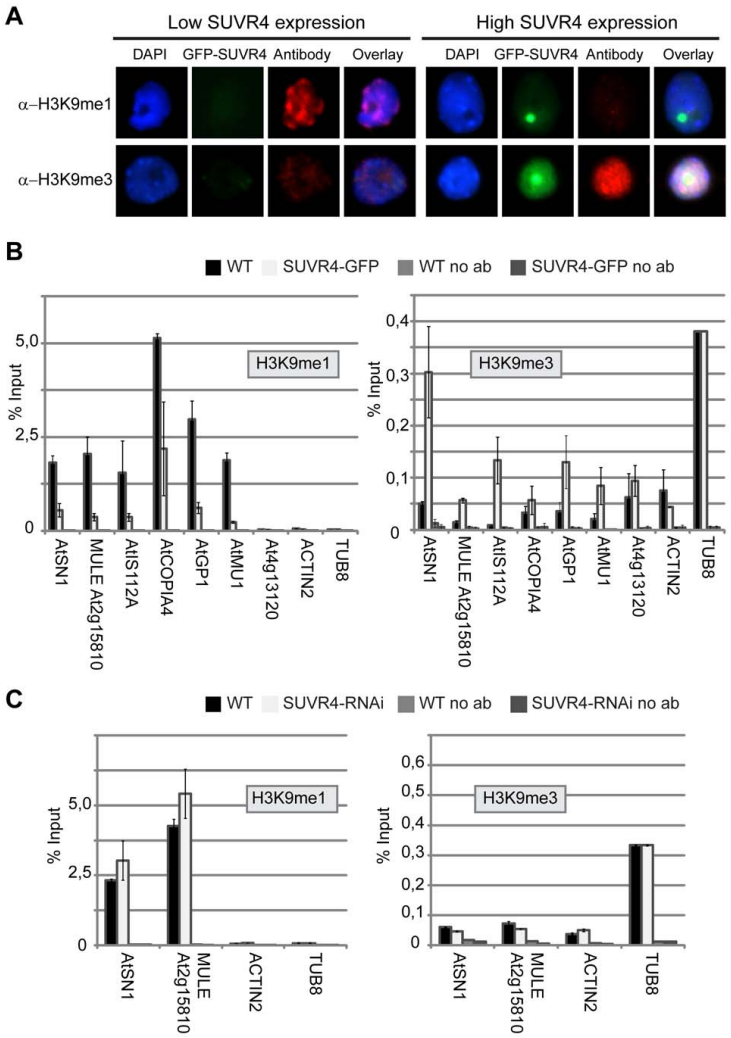

Figure 3. SUVR4 directs H3K9me3 on transposon and repeat sequences. (A) Immunostaining of nuclei from SUVR4-GFP^{OE} seedlings with low expression (left panel) or high expression (right panel) of SUVR4-GFP with antibodies against H3K9me1 or H3K9me3. ChIP analysis of (B) SUVR4-GFP^{OE} and (C) SUVR4 RNAi lines using antibodies against H3K9me1 (left) or H3K9me3 (right). DNA levels from the ChIP experiments (B, C) relative to the input reactions were quantified using real time PCR and normalized to *TUB8*. The bars represent the average of two independent biological replicates.
doi:10.1371/journal.pgen.1001325.g003

performed on two of the transposons that are targets of SUVR4 histone lysine methylation. We did not detect an effect of SUVR4 activity on DNA methylation of the *MULE At2g15810* transposon for CG, CHG or CHH in neither SUVR4 OE nor SUVR4 RNAi lines (Figure 5B). This suggests that the repressive effect of H3K9me3 added by SUVR4 is not mediated by DNA methylation. In contrast, the *AtSN1* transposon showed an increase in CHH methylation (Figure 5C) in the OE line. The CG and CHG methylation levels were unaffected. There was, however, no corresponding reduction of CHH methylation in the RNAi-line.

The ubiquitin protease UBP26 regulates the H3K9me2 and H3K9me3 level on transposons

The ubiquitin binding properties of the SUVR4 WIYLD domain and the ubiquitin-enhanced H3K9me3 activity of SUVR4 *in vitro* led us to look for links between ubiquitin and H3K9 trimethylation *in vivo*. Interestingly, deubiquitination of H2BUb1 by the nuclear UBP26/SUP32 ubiquitin protease, is required for repression of transposons [11], which also are targets of SUVR4. Therefore we investigated the H3K9me levels in the *ubp26-1/sup32* mutant (Figure S6). No effect was seen on highly expressed

Table 1. Transposon expression in various mutant backgrounds.

Gene/transposon	Agi Code	Type	Localization	*mom1*[a,b]	*kyp*[c,d]	*dc/ddc*[a]	*met1*[c,d]	K27 me3[c]
AtSN1	At3g44000/5	Retrotransposon	Euchromatin	-	X/UP	nd	X/UP	yes
MULE	At2g15810	DNA transposon	Euchromatin	UP	X/NoE	-	X/UP	yes
AtIS112A	At4g04293	DNA transposon	Euchromatin	UP	nd	UP	nd	nd
ATCOPIA4	At4g16870	Retrotransposon	nd	nd	X/Up	nd	X/UP	nd
ATGP1	At4g03650	Retrotransposon	Heterochromatin	nd	-	nd	X/UP	nd
AtMu1	At4g08680	DNA transposon	Heterochromatin	-	X/NoE	nd	X/UP	nd
AT4G13120	AT4g13120	DNA transposon	Euchromatin	nd	nd	nd	nd	nd
ACTIN2	AT3g18780	Non-TE control	Euchromatin	nd	-	-	-	nd
TUB8	AT5g23860	Non-TE control	Euchromatin	nd	-	-	-	yes

a) Numa et al., 2010 [52],
b) Habu et al.,2006 [33],
c) Mathieu O, Probst AV, Paszkowski J (2005) Distinct regulation of histone H3 methylation at lysines 27 and 9 by CpG methylation in Arabidopsis. EMBO J 24: 2783-2791
d) Lippman Z, Gendrel AV, Black M, Vaughn MW, Dedhia N, et al. (2004) Role of transposable elements in heterochromatin and epigenetic control. Nature 430: 471-476.
NoE = No expression, UP = increased transcription, X = affected in histone or DNA methylation, - = not affected, nd = not determined.
doi:10.1371/journal.pgen.1001325.t001

genes like *TUB8* and *ACTIN2* (Figure 6), and consistent with earlier findings [11], our ChIP analysis showed a reduction of H3K9me2 on transposons and repeat sequences (Figure 6A). Similarly, H3K9me3 was also reduced on transposons in the mutant compared to the wild type (Figure 6B). Although mutation in the *UBP26/SUP32* gene has been reported to lead to a global accumulation of H2Bub1 [35], the H2Bub1 level on transposons was only weakly affected by the mutation (Figure 6C), and the level of free ubiquitin monomers in the nuclei of *ubp26-1/sup32* was similar to the level in the wild type (Figure 6D).

We next tested the effect of global reduction of H2Bub1 on H3K9me3 level on transposon chromatin using the *hub2-2* mutant. This mutant is defect in the HISTONE MONOUBIQUITINATION2 E3 ligase, which acts non-redundantly with HUB1 to monoubiquitinate histone H2B [13]. The *hub2-2* mutant showed an almost complete lack of H2Bub1 at the *TUB8* gene, while the effect was absent or negligible on the *AtGP1* transposon. As reported for H3K9me2 [13,36], the H3K9me3 level was not affected either on *TUB8* or on transposon chromatin (Figure S7).

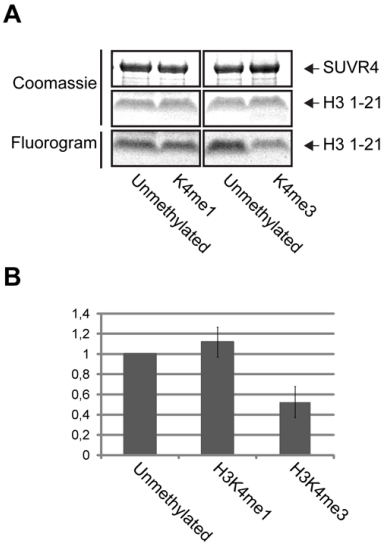

Figure 4. SUVR4 HKMTase activity is inhibited by H3K4me3. (A) HKMTase assay showing SUVR4 activity on peptides covering the first 1-21 aa of histone H3, that are unmodified, monomethylated or trimethylated on K4. (B) Quantification of band intensity from fluorogram in A, relative to the reaction with unmodified H3 1-21 peptide. The bars represent the average of three independent HKMTase assays.
doi:10.1371/journal.pgen.1001325.g004

Discussion

H3K9me3 has only recently been confirmed as a histone modification present in Arabidopsis, and its significance in gene regulation has only been indicative [23,24]. The presented work identifies SUVR4 as the first histone H3K9me3 methyltransferase in Arabidopsis and demonstrates how it cross-talks to ubiquitin and chromatin modifications like H3K9me1 and H3K4me3 to repress transposon transcription.

The WIYLD domain is a ubiquitin-binding domain pivotal for the HKMTase activity of the SUVR4 protein

Our experiments have identified the WIYLD domain of the SUVR4 HKMTase as a new ubiquitin interacting domain, demonstrating a direct link between ubiquitin binding and H3K9 methylation. Ubiquitin is extensively distributed in the eukaryotic proteome, and exists as free ubiquitin monomers, ubiquitin extension proteins, polyubiquitin, or ubiquitin conjugates [14]. The interactions with free ubiquitin, the ubiquitin moiety of the ubiquitin extension protein UBQ1 and the ubiquitin conjugate H2Bub1 (Figure 1), indicate that the SUVR4 WIYLD domain can target ubiquitin either in its free or conjugated form.

The interaction between the WIYLD domain of SUVR4 and ubiquitin is further supported by the WIYLD-dependent positive effect of ubiquitin on enzymatic activity (Figure 2). Free ubiquitin stimulated the HKMTase activity of the full-length SUVR4 protein without compromising the substrate specificity because no histones other than H3 were methylated (Figure 2C). However,

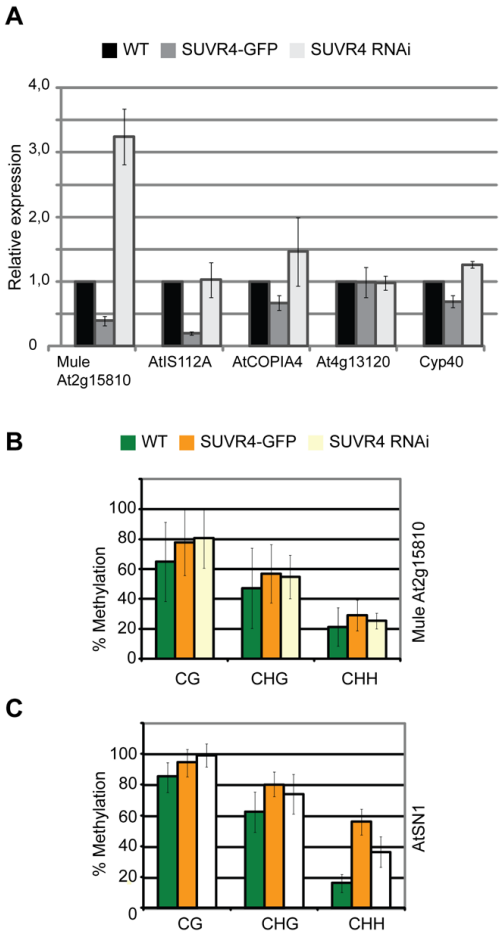

Figure 5. SUVR4 represses transcription of transposons. (A) Real time RT-PCR quantification of transcripts reversely transcribed from mRNA isolated from 14 day old SUVR4-GFP^{OE} and SUVR4-RNAi seedlings, respectively. The data were normalized to *ACTIN2* and shown relative to wild type. (B, C) Quantification of bisulfite treated DNA from wt, SUVR4^{OE} and SUVR4 RNAi seedlings for *MULE At2g15810* (B) and *AtSN1* (C) respectively.
doi:10.1371/journal.pgen.1001325.g005

the addition of free ubiquitin (Ub) converted the protein from a strict H3K9me2 to a H3K9me2/me3 methyltransferase (Figure 2D, 2E), suggesting that ubiquitin either in its free form or conjugated to other proteins like H2B can act as a signal for H3K9 trimethylation. We only observed 3% conversion of H3K9me1 to H3K9me3 after a 3 hour reaction time in our *in vitro* HKMTase assay while most of the H3K9me1 was converted to H3K9me2 (Figure 2E). In contrast, a massive shift from H3K9me1 to H3K9me3 was seen *in vivo* when over-expressing SUVR4 (Figure 3A, 3B). Together this implies the need for another component in addition to ubiquitin for SUVR4 to efficiently convert H3K9me1 to H3K9me3 *in vitro*, as shown for the murine ESET HKMTase [37]. In recombinant form *in vitro* ESET only catalyzes mono- and dimethylation of H3K9, but in complex with the transcriptional repressor mAM the enzyme generates H3K9me3.

Interestingly, the truncated SUVR4 SACSET protein showed a lower HKMTase activity compared to the full-length SUVR4 protein on core histones (Figure 2A), arguing that the N-terminal WIYLD domain is essential for normal activity of the C-terminal SET domain. Furthermore, the activity of the SUVR4 SACSET was only weakly enhanced by ubiquitin (Figure 2A, 2B), demonstrating that ubiquitin in its free form stimulates SUVR4 activity mainly through the WIYLD domain. Several enzymes that are involved in Ub pathways have shown to be regulated by ubiquitin. Recently, the activity of the mammalian deubiquitination enzyme ataxin-3 was shown to be enhanced by ubiquitination [38], and binding of free ubiquitin to the N-terminal ZnF-UBP domain of the deubiquitinase USP5 led to a conformational change that stimulated enzyme activity [31].

SUVR4 converts H3K9me1 to H3K9me2/me3 at transposons

In Arabidopsis H3K9me3 methylation broadly marks 40% of all genes within euchromatin [39]. In addition a low but detectable level of H3K9me3 methylation is found in regions with silenced transposons and pseudogenes [24] (Figure 3 and Figure 6). Our ChIP results suggest that although associated with both eu- and heterochromatin, SUVR4 has no HKMTase activity on euchromatic genes, but specifically targets transposons and repeat sequences where it converts H3K9me1 to H3K9me3 (Figure 3B, 3C). This is perfectly in line with our *in vitro* HKMTase results, which show that SUVR4 preferably uses H3K9me1 as substrate (Figure 2D). Together the *in vivo* and *in vitro* data indicate that SUVR4 only methylates transposons with a high H3K9me1 level although the protein might also associate with regions with a low level of this modification (Figure S3).

SUVR4 methylates unmethylated H3 poorly, and the level of H3K9me1 decreases in the OE line (Figure S1A and Figure 3B). This suggests that SUVR4 does not itself monomethylate H3K9 *in vivo*. Both SUVH4 and SUVH6 are efficient monomethyl transferases *in vitro* [40], which together with SUVH5 control the deposition of the majority of H3K9me1 at transposons and repeat sequences [41]. As SUVR4 targets the same type of sequences, it is likely that SUVR4 uses the monomethylated histone substrates created by the SUVH proteins to trimethylate H3K9. In mammalian cells, the SUV39H1 HKMTase depends on a monomethylase as it preferably converts H3K9me1 of H3.1, but not H3K9me2 of H3.3, to H3K9me3. [42]. Similarly, SUVR4 is stimulated by H3K9me1, but is only active on H3K9me2 if ubiquitin is added to the *in vitro* reaction.

The SUVH2 HKMTase has a strong impact on centromeric and pericentromeric heterochromatinization and gene silencing and reduces the level of H3K9me3 when overexpressed [32]. In contrast, overexpression of SUVR4 leads to increased H3K9me3 levels, and no changes in heterochromatinization could be observed (Figure 3A). Pericentromeric regions contain high levels of H3K9me1 and H3K9me2 in plants, but also H3S10 phoshporylation during mitosis and meiosis II [22]. The cell cycle dependent H3S10ph modification generated by Aurora kinase 1 inhibits SUVR4 activity *in vitro* [43]. This and the uninterrupted regions of high levels of H3K9me2 associated with the many transposons and pseudogenes located in pericentromeric and centromeric heterochromatin [44], may contribute to repress SUVR4 activity in these regions in dividing cells. Alternatively, SUVR4 might be able to methylate histones in pericentric heterochromatin before H3S10ph is added as Aurora kinase 1 is

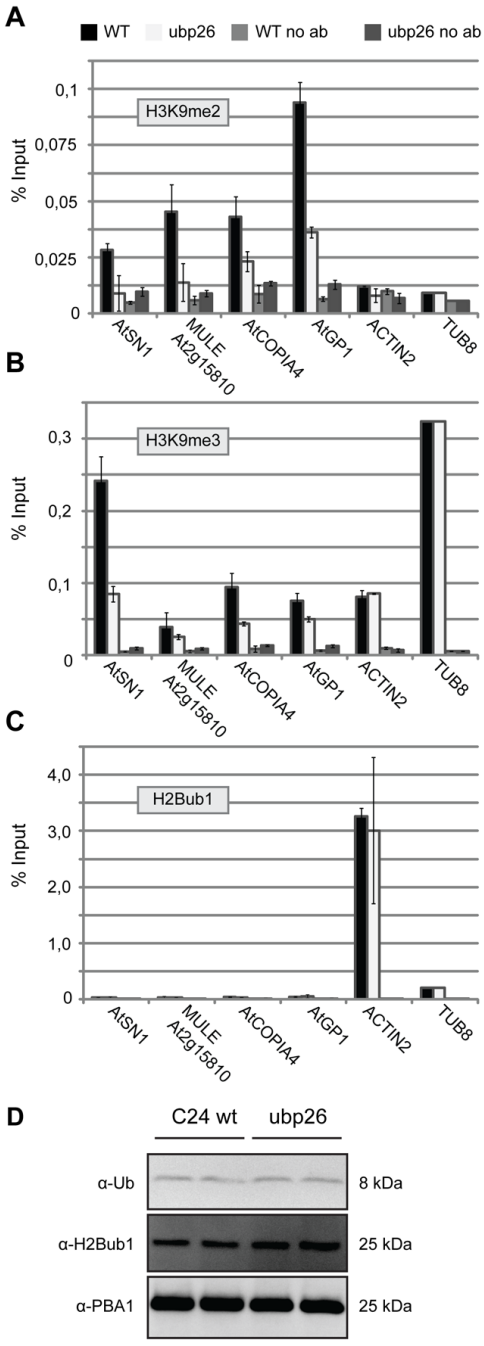

Figure 6. UBP26 directs H3K9me2 and H3K9me3 on transposon sequences. ChIP analysis of *ubp26-1* lines using antibodies against (A) H3K9me2, (B) H3K9me3 or (C) H2Bub1. DNA levels from the ChIP experiments relative to the input reactions were quantified using real time PCR and normalized to *TUB8*. The bars represent the average of two independent biological replicates. (D) Western blot of nuclear proteins isolated from *ubp26-1* and wild type, probed with antibodies against ubiquitin (ub), H2Bub1 or PBA1 (loading control).
doi:10.1371/journal.pgen.1001325.g006

active on methylated histones. Although pericentric heterochromatin most likely is not the preferred target of SUVR4 activity because of the high level of uninterrupted H3K9me2 [44], SUVR4 could potentially methylate transposons in these regions under certain conditions when ubiquitin levels are high, as demonstrated by the ability of SUVR4 to methylate H3K9me2 peptides when ubiquitin is added (Figure 2D, 2E, Figure S1B, and Figure 7B).

Mutation in the SUP32/UBP26 deubiquitinating enzyme that removes the ubiquitin conjugate from H2Bub1 has been reported to lead to reduction in H3K9me2 [11]. Using ChIP analysis we found low levels of H2Bub1 at all tested transposons, which were only weakly altered in the *ubp26* mutant line (Figure 6C). A reduction of both H3K9me2 and H3K9me3 was, however, observed on the same sequences targeted by SUVR4 (Figure 3B, 3C and Figure 6A, 6B). We therefore suggest that SUVR4 and UBP26 act in the same pathway leading to repression of transposon activity, and speculate that the reduction of H3K9me3 in *ubp26-1* mutant background can be due to reduced SUVR4 activity. Thus UBP26 can repress transposon transcription by lowering the H2Bub1 level at these sequences to maintain repressive H3 methylation as suggested by Sridhar et al. [11], and/or by maintaining a high local level of free ubiquitin which stimulates SUVR4-mediated H3K9me3 (Figure 7). Possibly UBP26/SUP32 can also cleave the ubiquitin extension protein UBQ1 initially found in our yeast two-hybrid screen to obtain free ubiquitin, as it has been shown to also be active on the human homologue CEP52 [11] which has 92% sequence identity with UBQ1. We did not however observe any reduction of free ubiquitin in the nuclear extracts of *ubp26-1* mutants (Figure 6D) that might have affected SUVR4 activity, and there was no effect on H3K9me3 or H2Bub1 at transposon sequences in the *hub2-2* line (Figure S7). Thus, HUB2 seems not to be involved in regulation of H2Bub1 or H3K9me2/3 or to be the counterpart of UBP26 on transposon chromatin. The minor reduction of H2Bub1 at transposons and the ability of UBP26/SUP32 to deubiquitinate the CEP52 *in vitro*, opens the possibility that UBP26 regulates SUVR4-dependent H3K9me2/3 by additional mechanisms, for instance transient changes in the levels or subnuclear distribution of free ubiquitin.

Highly transcribed euchromatic genes like *ACTIN2* and *TUB8* were unaffected by SUVR4, and the *in vitro* assay implies that SUVR4 activity is inhibited by H3K4me3 which is abundant in euchromatin (Figure 4). Furthermore, the *in vivo* data shows that the targets for SUVR4 activity have low levels of H3K4me3, H3K9me3 and H2Bub1 (Figure 3, Figure 6, and Figure S4). Intercalary heterochromatic sequences located within euchromatin are associated with intermediate amounts of opposing histone marks like H3K4me2 and H3K9me2 [33,44], but have comparable levels of H3K9me1 as heterochromatin (Figure 3B, 3C). As depicted in the model in Figure 7, this suggests that SUVR4 cross-talks to other PTMs and preferably targets transposons outside pericentric and centromeric heterochromatin, with low H3S10ph, H3K9me2, H3K4me3 and H2Bub1 and high H3K9me1 in order to trimethylate H3K9.

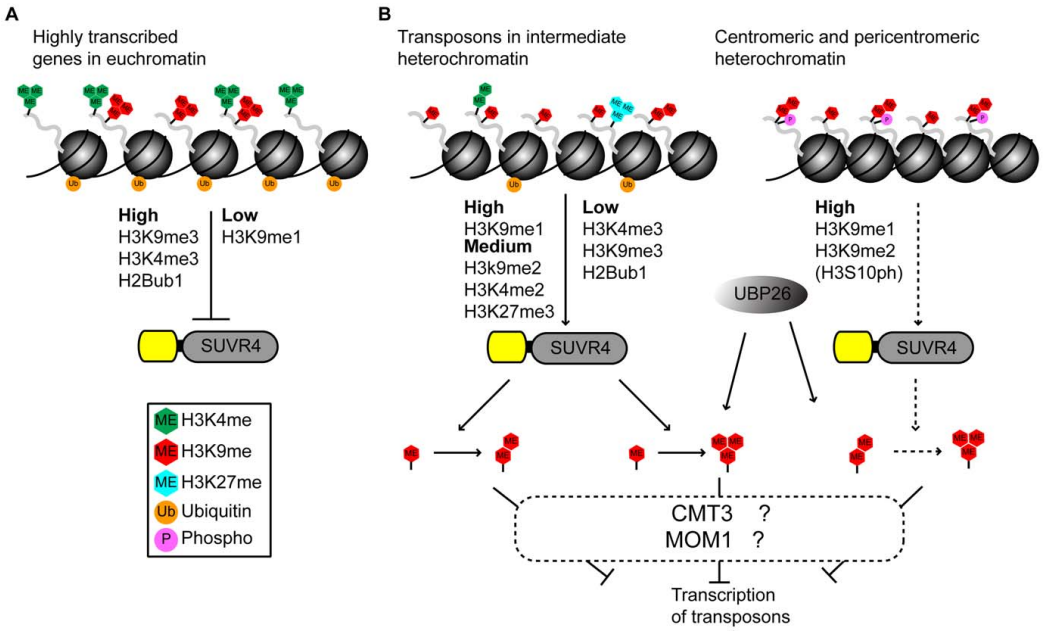

Figure 7. Model describing the relationship between free ubiquitin and SUVR4 activity on transposons. (A) SUVR4 is repressed by H3K4me3 *in vitro*, and has no activity on genes with high H3K4me3, H3K9me3, H2Bub1 and a low level of H3K9me1. (B) SUVR4's preference for heterochromatic transposons intercalated within euchromatin is maintained by its specificity for H3K9me1 which is highly enriched at transposons, and its repression by activating marks like H3K4me3. The deubiquitinase UBP26 regulates H3K9me2/me3 at the same targets as SUVR4, and might produce free ubiquitin that stimulates the H3K9me2/me3 activity of SUVR4 at target transposons. Although SUVR4 normally is repressed by H3K9me2 and H3S10ph which is high in pericentric heterochromatin, these regions may be targets for SUVR4 activity when ubiquitin levels are high. Since the transposons also contain a medium level of H3K27me3 in addition to H3K9me3, this could possibly create a binding site for CMT3 in order to repress transcription in a DNA methylation-dependent manner at some transposons. At other transposons, transcription may be repressed in a DNA methylation- independent manner by the MOM transcriptional repressor (See text for details).
doi:10.1371/journal.pgen.1001325.g007

SUVR4-mediated conversion of H3K9me1 to H3K9me3 represses transposon transcription in a locus specific manner

For transposon sequences with a low or intermediate expression level in wild type plants, increase in H3K9me3 levels mediated by SUVR4 is associated with repression of transcription (Figure 3, Figure 5, and Figure 7). In the RNAi line only the *MULE At2g15810* transposon, localized in euchromatin outside the typical pericentric heterochromatin or centromeric regions [33], showed relief of repression (Figure 5A), suggesting it to be a normal target of SUVR4 activity. However, *AtIS112A*, another transposon intercalated in euchromatin with an intermediate expression level, was only affected in the OE line. The heterochromatin localized *AtMU1* and the euchromatin localized *AtSN1*, both silent in wild type plants, were also targets for SUVR4 methylation but showed no reactivation in the RNAi line. This suggests that SUVR4-directed H3K9me3 regulates transposon activity in a locus specific manner, where SUVR4 activity alone is sufficient for repression of *MULE At2g15810*, while it works redundantly with an unknown HKMTase at other elements like *AtIS112A*, *AtMU1* and *AtSN1*. A similar regulation can be seen for the SUVH2 and SUVH9 SET domain proteins that act redundantly at some loci but independently at others [45]. Thus different transposons are regulated by different combinations of epigenetic marks (Table 1).

Genes in euchromatin have a much higher level of H3K9me3 than transposons, and in these regions this modification seems to correlate with activation of transcription and the deposition of other activating marks [23,24]. This argues for a combinatorial readout where the context of other PTMs with which H3K9me3 appears decides the biological outcome (Figure 7). In contrast to genes, transposon and repeat sequences contain a high level of H3K9me1 and low levels of H3K4me3 and H2Bub1 (Figure 3B, 3C, and Figure S4) and in this context H3K9me3 may lead to repression of transcription.

H3K9me1 on transposon chromatin seems to be a prerequisite and the preferred substrate for SUVR4 activity, as the control transposon At4g13120, with very low H3K9me1, was not methylated or affected at the transcriptional level (Figure 5A). Several studies have reported the accumulation of H3K9me1 in heterochromatin (reviewed in [22]) but little is known about the function of this mark. Our data supports a model where H3K9me1 is associated with both pericentric and centromeric heterochromatin and transposons intercalated in euchromatin, but does not act as a repressive signal, but rather a template for other methyltransferases. This is supported by the observation that increased H3K9me1 level correlated with increased transcription in the SUVR4 RNAi line and inversely correlated with increased H3K9me3 and repression of transcription in the SUVR4-GFP^{OE} line (Figure 3A–3C and Figure 5A).

H3K9me3 by SUVR4 may promote methylation-dependent and -independent repression of transposons

The level of DNA methylation of the *MULE At2g15810* transposon did not correlate with SUVR4 expression. At the *AtSN1* transposon, however, increased H3K9me3 mediated by SUVR4 overexpression coincided with an increase of CHH while no effect was seen for CG methylation (Figure 5B, 5C). Pericentric H3K9me2 shows a strong correlation with CHG methylation but a weaker correlation with CG and CHH methylation [44], while transposons located outside pericentric or centromeric heterochromatin have shorter patches of H3K9me2 at lower levels. Together with the repressive effect of H3K9me2 on SUVR4 activity this argues that the main DNA methylation regulated by SUVR4 is CHH.

The DRM2 methylase is the main regulator of asymmetric CHH methylation, while CHROMOMETHYLASE3 (CMT3) is the main regulator of CHG methylation in Arabidopsis, but at some loci they work together [46,47]. At dispersed repeats within euchromatin like *AtSN1*, DRM1, DRM2 and CMT3 act redundantly to maintain CHH and CHG methylation [48]. At such loci we suggest that the H3K9me3 methylation by SUVR4 might mark the underlying transposon sequence for CHH methylation by DRM2/CMT3 (Figure 7B). Interestingly, many transposon sequences contain both H3K27me3 and H3K9me3, a combination that CMT3 has been shown to bind *in vitro* (Table 1, [24,30,49,50]). The redundant regulation of *AtSN1* by CMT3 and DRM1 might thus explain the lack of reactivation and DNA methylation upon reduction of SUVR4 H3K9me3 methylation in the SUVR4 RNAi line.

Although a target of SUVR4-directed H3K9me3 and repression, the *MULE* transposon was not affected at the DNA methylation level (Figure 5B). In contrast to *AtSN1*, this transposon has been shown earlier to be activated only in *mom1* mutants, and not in mutants with reduced non-CG methylation and *kyp/suvh4* mutants (Table 1). MOM1 is a transcriptional repressor that regulates transcriptional gene silencing of loci outside centromeric and pericentromeric heterochromatin, with only small effects on epigenetic marks [33,51,52]. This suggests that non-CG methylation is not involved in silencing of *MULE*. The similar relief of silencing without any effect on DNA methylation between SUVR4 RNAi and *mom1* makes it tempting to speculate that SUVR4 recruits MOM1 to its targets in order to repress transcription at this locus (Figure 7B). The intermediately expressed *AtIS112A* is repressed in SUVR4 OE lines but did not show any relief of expression in the RNAi line. As for *AtSN1*, this transposon is regulated by non-CG methylation, but also by MOM1. This argues that SUVR4 mediated repression might act via DNA methylation-independent mechanisms such as for *MULE At2g15810*, but also by DNA methylation-dependent mechanisms as seen for *AtSN1*, or possibly both as seen for *AtIS112A*.

DUBs are important to maintain ubiquitin homeostasis by recycling ubiquitin from free ubiquitin chains, ubiquitin conjugates and ubiquitin fusion proteins [14,15]. UBP26 regulates H3K9me2 and H3K9me3 methylation as well as non-CG methylation at the same sequences as SUVR4 [11]. We hypothesize that UBP26 acts in concert with SUVR4 to trimethylate transposons with a high level of H3K9me1 and low level of H3K4me3 and H2Bub1 (Figure 7). The H3K9me3 methylation thus directs locus-specific methylation-dependent or -independent repression of transposon activity.

Methods

Plant material

Arabidopsis plants, ecotype Columbia (Col), were grown under long day greenhouse conditions at 18°C. Transgenic *Arabidopsis* plants were generated by the floral dip method [53] using the *Agrobacterium tumefaciens* strain C58 pCV2260. Transgenic plants containing the pEG104 [54] or pART27 [55] vectors were selected on MS-2 medium (1x Murashige and Skoog salts, 0.05% 2-N-morpholino/ethanesulfonic acid, 2% sucrose, 0.8% agar) containing 10 µg/ml basta or 50 µg/ml kanamycin, respectively. For ChIP, RT-PCR and cytology experiments, Col wild type plants and non-segregating lines containing the respective T-DNA constructs were grown on MS-2 without antibiotic selection. The *ubp26*-mutant [11] and the *hub2-2* [13] mutant lines have been described earlier.

RNA extraction, cDNA synthesis, and real-time PCR

RNA was isolated from approx. 100 mg of 14 day old seedlings using the Spectrum Plant Total RNA Kit with on-column DNase treatment (Sigma). cDNA synthesis and Real time RT-PCR experiments were performed as described previously [20] using gene specific primers (Table S1), except that 4 µg of total RNA was used to synthesize first strand cDNA with Superscript III Reverse Transcriptase and random primers (Invitrogen).

DNA constructs

SUVR4-Full (At3g04380), SUVR4-SACSET, SUVR4-WIYLD, UBQ1, ubiquitin moiety of UBQ1 and L40 moiety of UBQ1 were PCR amplified from cDNA using gene specific *attB* gateway primers (Table S1) and *Pfu* DNA polymerase (Fermentas). The attB PCR products were recombined into the pDONR/Zeo vector using the Gateway BP Clonase II Enzyme Mix (Invitrogen) according to the manufacturer's instructions. The resulting pDONR/Zeo entry clones were recombined into destination vectors using the Gateway LR Clonase Enzyme Mix (Invitrogen). All constructs were verified by sequencing. The knock-down SUVR4 RNAi construct was made by cloning a unique fragment from the SUVR4 5′end as an inverted repeat on each side of an intron into the binary vector pART27. Cloning procedures are described in detail (Text S1).

Yeast two-hybrid screening

Two-hybrid interactions were screened by mating the yeast strain Y187 carrying the pGBKT7-SUVR4-WIYLD bait construct with the strain AH109 carrying a cDNA library (Matchmaker library construction and screening kit, Clontech) at 30°C ON. The cDNA library was created from Columbia wt 14 day old seedlings and recombined into the pGADT7-Rec vector to create an AD-fusion library. Selective media for the nutritional reporter genes ADE2, HIS3 and MEL1 (QDO) containing 20 mg l-1 X-alpha-Gal, was used to identify positive two-hybrid interactions according to the suppliers suggestions. To confirm interaction with SUVR4-WIYLD, the pGADT7-UBQ1, pGADT7-ubiquitin and pGADT7-L40 were mated separately with the pGBKT7-SUVR4-WIYLD or the empty pGBKT7 vector (BD control). Diploid colonies were selected on SD –L/-T, and then streaked out on SD –L/-T/-H +3 AT medium selective for protein-protein interactions.

Expression of recombinant proteins for enzyme assays

pHMGWA-SUVR4-Full and pHMGWA-SUVR4-SACSET constructs were transformed into *E. coli* BL21-Star DE3 and grown at 150 rpm, 37°C in LB-medium with 1% Glucose and 100 µg/ml ampicillin. At an OD_{600} 0.6–0.8, the cells were induced with 1 mM IPTG overnight at 20°C. The cells were lysed with Express and then resuspended in pre-cooled lysis Buffer: 20 mM Tris–HCl, pH 7.5, 400 mM NaCl, 100 mM KCl, 1 mM

EDTA, 1 mM DTT, 0.05% Triton X-100 and Protease inhibitor. After centrifugation (15,000 rpm), the supernatant containing recombinant protein was filtered through 0.45 µm filters and prepared for affinity chromatography.

Purification of recombinant proteins

Recombinant proteins SUVR4-Full and SUVR4-SACSET were purified by Ni-NTA affinity chromatography using HisTrap FF 5 ml (GE Healthcare) column in the ÄKTA purifier. Binding buffer or Buffer A and Elution Buffer or Buffer B in the purification step were as follows, Buffer A: 20mM Tris–HCl, pH 7.5, 500mM NaCl, 1 mM EDTA, 1 mM DTT, 20 mM Imidazole and Buffer B: 20 mM Tris–HCl, pH 7.5, 500 mM NaCl, 1 mM EDTA, 1 mM DTT, 500 mM Imidazole.

HKMTase assays

HKMTase assays were essentially performed as described in [28]. Twenty µg of MBP-SUVR4 protein was incubated in reaction buffer (50 mM Tris pH 8.5, 20 mM KCl, 20 mM $MgCl_2$, 10 mM β-mercaptoethanol and 250 mM sucrose) with 7.5 µl µCi 14C S-adenosyl methionine (SAM) (Amersham/Perkin Elmer) or 100 µM unlabelled SAM (New England Biolabs) as methyl donor. Twenty µg of core histones from calf thymus (Roche), or 5 µg histone H3 peptides were used as substrate. Reactions were incubated at 30°C for 3 hours, and each experiment was repeated at least 4 times. Core histones from calf thymus (Roche), unmodified histone H3 peptide (#12-403, Millipore), monomethyl-histone H3 (Lys9) peptide (#12-569, Millipore), dimethyl-Histone H3 (Lys9) Peptide (#12-430, Millipore), Trimethyl-Histone H3 (Lys9) Peptide (#12-568, Millipore), Trimethyl-Histone H3 (Lys4) Peptide (#12-564, Millipore), monomethyl histone H3 (Lys 4) peptide (gift from Thomas Jenuwein) and ubiquitin (U6253, Sigma) were used in the assays.

GST pull-down

Recombinant proteins were expressed in BL21 cells, lysed in 1 X PBS with 0.1 mg/ml lysozyme, 0.2–1% Triton X-100 and protease inhibitor cocktail (Roche), and immobilized on glutathione sepharose beads (Amersham). 3 µg of GST-S4WIYLD was incubated with MBP protein lysates at 4°C for 2.5 hours or 10 µg of GST-SUVR4-WIYLD with 20 µg of precleared core histones (Roche) at 4°C for 3 hours, following a series of washes. Pull-down reactions were run on SDS-PAGE gels, blotted onto a PVDF membrane (Machery Nagel) and probed with either anti-MBP (1:10000, New England Biolabs, #E8030S) or anti-H2Bub1 (1:1000, MediMabs, MM-0029). Detection of primary antibody was performed with peroxidase-conjugated secondary antibody; goat anti-rabbit HRP for pulldown of MBP-proteins (1:10000, Thermo Scientific, PA1-74361) and anti-mouse HRP for pull-down of core histones (1:10000, Abcam, ab6728) using the ECL kit (GE HealthCare, RPN2135).

MS analysis of peptides from HKMTase reaction mixtures

Reverse phase (C18) nano online liquid chromatographic MS/MS analyses of proteolytic peptides from HKMTase reactions using unlabelled SAM were performed using a HPLC system as described [56].

Nuclear magnetic resonance spectroscopy

Uniformly ^{15}N- or ^{15}N,^{13}C-labeled SUVR4-WIYLD (residues 1-89) was expressed as a GST-fusion (pGEX4T3) in minimal media containing $^{15}NH_4Cl$ and ^{13}C-glucose as the sole nitrogen and carbon sources, respectively, after induction at 18°C for 18 hours. Protein was purified by glutathione sepharose affinity and size-exclusion chromatography and thrombin digestion to remove the affinity tag. NMR samples contained 0.5 mM protein in PBS at pH 7.4, 5 mM d_{10}-DTT and 10% D_2O. All spectra were acquired at 25°C on a 500MHz or 600MHz Bruker spectrometer.

Chromatin immunoprecipitation

For each experiment 2-3 g of fifteen day old seedlings was crosslinked in 1% formaldehyde under vacuum until the tissue was translucent. Chromatin immunoprecipitation was done as described in [57]. The antibodies used for immunoprecipitation were anti-H2Bub1 (#MM-0029, Medimabs), anti-H3K9me1 (#07-450, Millipore), anti-H3K9me2 (#07-212, Millipore) anti-H3K9me3 (#07-442, Millipore), anti-H3K4me3 (#07-473, Millipore) and anti-GFP (#ab290-50, Abcam). Immunoprecipitated chromatin was eluted in a total of 250 µl elution buffer (1% SDS, 0.1 M NaHCO3) and after reversion of crosslinking, DNA was extracted using the Qiaquick PCR purification kit (Qiagen) and eluted in 100 µl elution buffer. 5 µl of a 4 X dilution was used as a template for real-time PCR in a Lightcycler (Roche). Typically a program of: 1 cycle 95°C 10 min, 45 cycles of 95°C 20 s, 52° 30 s and 72°C 30 s was used to amplify target sequences with gene specific primers (Table S1). PCR was performed on ChIP DNA isolated from two independent experiments, each quantified two separate times.

Western blotting

Nuclear protein extracts were isolated from a chromatin preparation as described [57]. The protein lysate obtained after sonication was separated on a 10-20% SDS-PAGE (Invitrogen, catalog no. EC6625BOX) and transferred to a PVDF membrane (Machery Nagel). Nuclear protein levels were determined using the following antibodies; anti-ubiquitin (1:4000, Millipore, 07-375), anti-H2Bub1 (1:1000, MediMabs, MM-0029) and anti-PBA1 (1:1000, abcam, ab98999).

Immunostaining of nuclei

Leaves from 14 day old seedlings were chopped in 4% formaldehyde on slides, covered with coverslips and flash frozen in liquid N2. The coverslips were removed from the slides when the material was still frozen, and then the slides were washed three times 5 minutes in 1 X PBS. The material was then blocked for 30 min at 37°C in blocking solution (1% BSA in PBS), and incubated with primary antibody (anti H3K9me1, 1:200; antiH3K9me3, 1:100) diluted in blocking solution for one hour at 37°C. After a series of washes in PBS, the slides were incubated with goat-anti rabbit Alexa 555 (Invitrogen) secondary antibody (1:200). Before microscopy the slides were washed in PBS and counterstained in DAPI and inspected with a Zeiss Axiovision2 microscope equipped with epifluorescence attachment. All images were captured using the same exposure times and at 100X magnification.

Bisulfite sequencing

2 µg of genomic DNA, prepared from leaf material using the Invisorb Spin Plant Kit (INVITEK Berlin), was restricted with ApaI and PstI and used in the bisulfite reaction with the EpiTect Bisulfite Kit (Quiagene Hilden). Bisulfite treated DNA was used as template in a PCR with specific primers. The PCR-Fragments are ligated into pGEMT-vector (Promega) and transformed in DH5alpha cells. Plasmid DNA from several colonies was sequenced with the ABI Prism 310.

Supporting Information

Figure S1 HKMTase activity of SUVR4. (A) HKMTase assay with MBP-SUVR4 full-length using unmethylated histone H3 1-21 or histone H3 K9me1 peptides as substrate. (B) Second independent replica of the HKMTase assay in Fig. 2 D. MBP-SUVR4 full-length activity on histone H3 1-21 K9me2 peptides without and with the addition of 5 µg of free ubiquitin. (C) HKMTase assay with MBP-SUVR4 full length, MBP-SUVR4-SACSET, MBP-SUVR4-WIYLD and no protein on core histones without and with the addition of 5 µg of free ubiquitin.
Found at: doi:10.1371/journal.pgen.1001325.s001 (2.43 MB TIF)

Figure S2 SUVR4 subcellular localization. Fluorescence microscopy of interphase nuclei from seedlings expressing SUVR4-GFP fusion proteins, demonstrating varying subcellular localization. (A) Nucleus showing uniform SUVR4 localization to the nucleoplasm and nucleolus (no), with high accumulation in an unknown focus (uf). (B) Uniform SUVR4-GFP distribution in the nucleoplasm, with strong localization in nucleolar associated foci and weaker localization to the nucleolus. (C) Strong SUVR4 localization to the nucleolus and weaker association to the nucleoplasm. (D) SUVR4 localization to the nucleoplasm, with stronger accumulation in the nucleolus and an unknown focus.
Found at: doi:10.1371/journal.pgen.1001325.s002 (0.99 MB TIF)

Figure S3 SUVR4-GFP associates with eu- and heterochromatin. ChIP analysis of SUVR4-GFP^{OE} lines using an antibody against GFP. DNA levels from the ChIP experiments relative to the input reactions were quantified using real time PCR and normalized to *TUB8*. The bars represent the average of two independent biological replicates.
Found at: doi:10.1371/journal.pgen.1001325.s003 (0.34 MB TIF)

Figure S4 H2Bub1 levels on transposons. (A) ChIP analysis of SUVR4-GFP^{OE} lines using antibodies against H2Bub1. (B) ChIP analysis of SUVR4 RNAi lines using antibodies against H3K4me3. DNA levels from the ChIP experiments relative to the input reactions were quantified using real time PCR and normalized to TUB8. The bars represent the average of two independent biological replicates.
Found at: doi:10.1371/journal.pgen.1001325.s004 (0.86 MB TIF)

Figure S5 Expression levels of *SUVR4* and transpsons. (A) Real time RT-PCR quantification of transcripts reversely transcribed from mRNA isolated from 14 day old SUVR4-RNAi seedlings, using *SUVR4* primers. The data were normalized to *ACTIN2* and the mutant expression is relative to wild type. Error bars represent standard deviation according to three biological replicates (n = 3). (B) Real time quantification of transposon expression in wild type. The expression of each transposon is relative to *ACTIN2* which is set to 1. Reactions without the addition of reverse transcriptase (-RT) is used as a negative control.
Found at: doi:10.1371/journal.pgen.1001325.s005 (0.17 MB TIF)

Figure S6 Genotyping of the *ubp26*-mutant. PCR on wild type and ubp26 mutant plants using the primer combinations P1 (*ubp26-1* F) primer with P2 (*ubp26-1* R), or P1 (*ubp26-1* F) with LB, on two biological replicas b1 and b2 (upper panel). Layout of the the *ubp26* gene indicating the position of the T-DNA insertion and the primer annealing sites (lower panel).
Found at: doi:10.1371/journal.pgen.1001325.s006 (1.20 MB TIF)

Figure S7 ChIP analysis of *hub2-2* plants. ChIP analysis of *hub2-2* and wild type plants using antibodies against H3K9me3 (A) or H2Bub1 (B). DNA levels from the ChIP experiments relative to the input reactions were quantified using real time PCR and normalized to *TUB8*. The data for H2Bub1 is not normalized to *TUB8* because the chromatin at this gene is affected by the *hub2-2* mutation. The bars represent the average of two independent biological replicates.
Found at: doi:10.1371/journal.pgen.1001325.s007 (0.64 MB TIF)

Table S1 Oligos used in this study.
Found at: doi:10.1371/journal.pgen.1001325.s008 (0.03 MB XLS)

Text S1 Cloning of DNA constructs.
Found at: doi:10.1371/journal.pgen.1001325.s009 (0.04 MB DOC)

Acknowledgments

We appreciate the technical assistance performed by Roy Falleth and Solveig H. Engebretsen. The work was facilitated by the services provided by the Norwegian Arabidopsis Research Centre (NARC, http://www.narc.no/), a part of the Research Council of Norway's National Program for Research in Functional Genomics (FUGE). We thank Thomas Jenuwein for the H3 1-21 K4me1 and H3 1-21 unmodified peptides, Rick Amasino for the *ubp26-1* seeds, Wim Soppe for the *hub2-2* seeds, and Sylvia S. Johnsen for the cloning of the SUVR4 RNAi construct.

Author Contributions

Conceived and designed the experiments: KLY AF WEJ TT. Performed the experiments: SVV MAR KLY AF WEJ TT. Analyzed the data: SVV MAR KLY AF WEJ TT. Contributed reagents/materials/analysis tools: GR MMZ RBA. Wrote the paper: TT. Contributed to paper writing: SVV KLY GR MMZ RBA.

References

1. Kouzarides T (2007) Chromatin modifications and their function. Cell 128: 693–705.
2. Latham JA, Dent SY (2007) Cross-regulation of histone modifications. Nat Struct Mol Biol 14: 1017–1024.
3. Taverna SD, Li H, Ruthenburg AJ, Allis CD, Patel DJ (2007) How chromatin-binding modules interpret histone modifications: lessons from professional pocket pickers. Nat Struct Mol Biol 14: 1025–1040.
4. Yap KL, Zhou MM (2006) Structure and function of protein modules in chromatin biology. Results Probl Cell Differ 41: 1–23.
5. de la Paz Sanchez M, Gutierrez C (2009) Arabidopsis ORC1 is a PHD-containing H3K4me3 effector that regulates transcription. Proc Natl Acad Sci U S A 106: 2065–2070.
6. Dikic I, Wakatsuki S, Walters KJ (2009) Ubiquitin-binding domains - from structures to functions. Nat Rev Mol Cell Biol 10: 659–671.
7. Hicke L, Schubert HL, Hill CP (2005) Ubiquitin-binding domains. Nat Rev Mol Cell Biol 6: 610–621.
8. Haglund K, Dikic I (2005) Ubiquitylation and cell signaling. EMBO J 24: 3353–3359.
9. Shukla A, Chaurasia P, Bhaumik SR (2009) Histone methylation and ubiquitination with their cross-talk and roles in gene expression and stability. Cell Mol Life Sci 66: 1419–1433.
10. Weake VM, Workman JL (2008) Histone ubiquitination: triggering gene activity. Mol Cell 29: 653–663.
11. Sridhar VV, Kapoor A, Zhang K, Zhu J, Zhou T, et al. (2007) Control of DNA methylation and heterochromatic silencing by histone H2B deubiquitination. Nature 447: 735–738.
12. Gu X, Jiang D, Wang Y, Bachmair A, He Y (2009) Repression of the floral transition via histone H2B monoubiquitination. Plant J 57: 522–533.
13. Cao Y, Dai Y, Cui S, Ma L (2008) Histone H2B monoubiquitination in the chromatin of FLOWERING LOCUS C regulates flowering time in Arabidopsis. Plant Cell 20: 2586–2602.
14. Komander D, Clague MJ, Urbe S (2009) Breaking the chains: structure and function of the deubiquitinases. Nat Rev Mol Cell Biol 10: 550–563.
15. Kimura Y, Yashiroda H, Kudo T, Koitabashi S, Murata S, et al. (2009) An inhibitor of a deubiquitinating enzyme regulates ubiquitin homeostasis. Cell 137: 549–559.
16. Lachner M, Jenuwein T (2002) The many faces of histone lysine methylation. Curr Opin Cell Biol 14: 286–298.
17. Berger SL (2007) The complex language of chromatin regulation during transcription. Nature 447: 407–412.
18. Fischer A, Hofmann I, Naumann K, Reuter G (2006) Heterochromatin proteins and the control of heterochromatic gene silencing in Arabidopsis. J Plant Physiol 163: 358–368.

19. Xu L, Zhao Z, Dong A, Soubigou-Taconnat L, Renou JP, et al. (2008) Di- and tri- but not monomethylation on histone H3 lysine 36 marks active transcription of genes involved in flowering time regulation and other processes in Arabidopsis thaliana. Mol Cell Biol 28: 1348–1360.
20. Grini PE, Thorstensen T, Alm V, Vizcay-Barrena G, Windju SS, et al. (2009) The ASH1 HOMOLOG 2 (ASHH2) histone H3 methyltransferase is required for ovule and anther development in Arabidopsis. PLoS ONE 4: e7817. doi:10.1371/journal.pone.0007817.
21. Fransz P, ten Hoopen R, Tessadori F (2006) Composition and formation of heterochromatin in Arabidopsis thaliana. Chromosome Res 14: 71–82.
22. Fuchs J, Demidov D, Houben A, Schubert I (2006) Chromosomal histone modification patterns—from conservation to diversity. Trends Plant Sci 11: 199–208.
23. Caro E, Castellano MM, Gutierrez C (2007) A chromatin link that couples cell division to root epidermis patterning in Arabidopsis. Nature 447: 213–217.
24. Charron JB, He H, Elling AA, Deng XW (2009) Dynamic landscapes of four histone modifications during deetiolation in Arabidopsis. Plant Cell 21: 3732–3748.
25. Baumbusch LO, Thorstensen T, Krauss V, Fischer A, Naumann K, et al. (2001) The Arabidopsis thaliana genome contains at least 29 active genes encoding SET domain proteins that can be assigned to four evolutionarily conserved classes. Nucleic Acids Research 29: 4319–4333.
26. Johnson LM, Bostick M, Zhang X, Kraft E, Henderson I, et al. (2007) The SRA methyl-cytosine-binding domain links DNA and histone methylation. Curr Biol 17: 379–384.
27. Liu C, Lu F, Cui X, Cao X (2010) Histone methylation in higher plants. Annu Rev Plant Biol 61: 395–420.
28. Thorstensen T, Fischer A, Sandvik SV, Johnsen SS, Grini PE, et al. (2006) The Arabidopsis SUVR4 protein is a nucleolar histone methyltransferase with preference for monomethylated H3K9. Nucleic Acids Res 34: 5461–5470.
29. Catic A, Ploegh HL (2005) Ubiquitin—conserved protein or selfish gene? Trends Biochem Sci 30: 600–604.
30. Zhang X, Clarenz O, Cokus S, Bernatavichute YV, Pellegrini M, et al. (2007) Whole-genome analysis of histone H3 lysine 27 trimethylation in Arabidopsis. PLoS Biol 5: e129. doi:10.1371/journal.pbio.0050129.
31. Reyes-Turcu FE, Horton JR, Mullally JE, Heroux A, Cheng X, et al. (2006) The ubiquitin binding domain ZnF UBP recognizes the C-terminal diglycine motif of unanchored ubiquitin. Cell 124: 1197–1208.
32. Naumann K, Fischer A, Hofmann I, Krauss V, Phalke S, et al. (2005) Pivotal role of AtSUVH2 in heterochromatic histone methylation and gene silencing in Arabidopsis. EMBO J 24: 1418-1429. Epub 2005 Mar 1417.
33. Habu Y, Mathieu O, Tariq M, Probst AV, Smathajitt C, et al. (2006) Epigenetic regulation of transcription in intermediate heterochromatin. EMBO Rep 7: 1279–1284.
34. Stancheva I (2005) Caught in conspiracy: cooperation between DNA methylation and histone H3K9 methylation in the establishment and maintenance of heterochromatin. Biochem Cell Biol 83: 385–395.
35. Schmitz RJ, Tamada Y, Doyle MR, Zhang X, Amasino RM (2009) Histone H2B deubiquitination is required for transcriptional activation of FLOWERING LOCUS C and for proper control of flowering in Arabidopsis. Plant Physiol 149: 1196–1204.
36. Dhawan R, Luo H, Foerster AM, Abuqamar S, Du HN, et al. (2009) HISTONE MONOUBIQUITINATION1 interacts with a subunit of the mediator complex and regulates defense against necrotrophic fungal pathogens in Arabidopsis. Plant Cell 21: 1000–1019.
37. Wang H, An W, Cao R, Xia L, Erdjument-Bromage H, et al. (2003) mAM facilitates conversion by ESET of dimethyl to trimethyl lysine 9 of histone H3 to cause transcriptional repression. Mol Cell 12: 475–487.
38. Todi SV, Winborn BJ, Scaglione KM, Blount JR, Travis SM, et al. (2009) Ubiquitination directly enhances activity of the deubiquitinating enzyme ataxin-3. EMBO J 28: 372–382.
39. Roudier F, Teixeira FK, Colot V (2009) Chromatin indexing in Arabidopsis: an epigenomic tale of tails and more. Trends Genet 25: 511–517.
40. Jackson JP, Johnson L, Jasencakova Z, Zhang X, PerezBurgos L, et al. (2004) Dimethylation of histone H3 lysine 9 is a critical mark for DNA methylation and gene silencing in Arabidopsis thaliana. Chromosoma 112: 308–315.
41. Ebbs ML, Bender J (2006) Locus-specific control of DNA methylation by the Arabidopsis SUVH5 histone methyltransferase. Plant Cell 18: 1166–1176.
42. Loyola A, Bonaldi T, Roche D, Imhof A, Almouzni G (2006) PTMs on H3 variants before chromatin assembly potentiate their final epigenetic state. Mol Cell 24: 309–316.
43. Demidov D, Hesse S, Tewes A, Rutten T, Fuchs J, et al. (2009) Aurora1 phosphorylation activity on histone H3 and its cross-talk with other post-translational histone modifications in Arabidopsis. Plant J 59: 221–230.
44. Bernatavichute YV, Zhang X, Cokus S, Pellegrini M, Jacobsen SE (2008) Genome-wide association of histone H3 lysine nine methylation with CHG DNA methylation in Arabidopsis thaliana. PLoS ONE 3: e3156. doi:10.1371/journal.pone.0003156.
45. Johnson LM, Law JA, Khattar A, Henderson IR, Jacobsen SE (2008) SRA-domain proteins required for DRM2-mediated de novo DNA methylation. PLoS Genet 4: e1000280. doi:10.1371/journal.pgen.1000280.
46. Cao X, Jacobsen SE (2002) Locus-specific control of asymmetric and CpNpG methylation by the DRM and CMT3 methyltransferase genes. Proc Natl Acad Sci U S A 99 Suppl 4: 16491–16498.
47. Henderson IR, Jacobsen SE (2007) Epigenetic inheritance in plants. Nature 447: 418–424.
48. Chan SW, Henderson IR, Jacobsen SE (2005) Gardening the genome: DNA methylation in Arabidopsis thaliana. Nat Rev Genet 6: 351–360.
49. Lindroth AM, Shultis D, Jasencakova Z, Fuchs J, Johnson L, et al. (2004) Dual histone H3 methylation marks at lysines 9 and 27 required for interaction with CHROMOMETHYLASE3. EMBO J 23: 4286–4296.
50. Law JA, Jacobsen SE (2010) Establishing, maintaining and modifying DNA methylation patterns in plants and animals. Nat Rev Genet 11: 204–220.
51. Yokthongwattana C, Bucher E, Caikovski M, Vaillant I, Nicolet J, et al. (2010) MOM1 and Pol-IV/V interactions regulate the intensity and specificity of transcriptional gene silencing. EMBO J 29: 340–351.
52. Numa H, Kim JM, Matsui A, Kurihara Y, Morosawa T, et al. (2010) Transduction of RNA-directed DNA methylation signals to repressive histone marks in Arabidopsis thaliana. EMBO J 29: 352–362.
53. Clough SJ, Bent AF (1998) Floral dip: a simplified method for Agrobacterium-mediated transformation of Arabidopsis thaliana. Plant J 16: 735–743.
54. Earley KW, Haag JR, Pontes O, Opper K, Juehne T, et al. (2006) Gateway-compatible vectors for plant functional genomics and proteomics. Plant J 45: 616–629.
55. Gleave AP (1992) A versatile binary vector system with a T-DNA organisational structure conducive to efficient integration of cloned DNA into the plant genome. Plant Mol Biol 20: 1203–1207.
56. Borud B, Aas FE, Vik A, Winther-Larsen HC, Egge-Jacobsen W, et al. (2010) Genetic, structural, and antigenic analyses of glycan diversity in the O-linked protein glycosylation systems of human Neisseria species. J Bacteriol 192: 2816–2829.
57. Gendrel A-V, Lippman Z, Martienssen R, Colot V (2005) Profiling histone modification patterns in plants using genomic tiling microarrays. Nature Methods 2: 213–218.

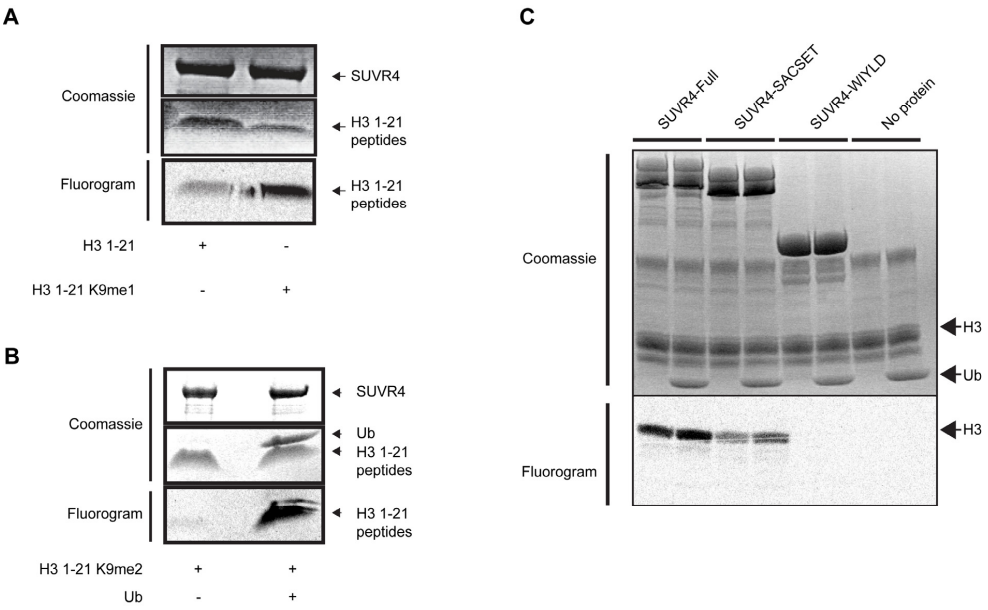

Figure S1

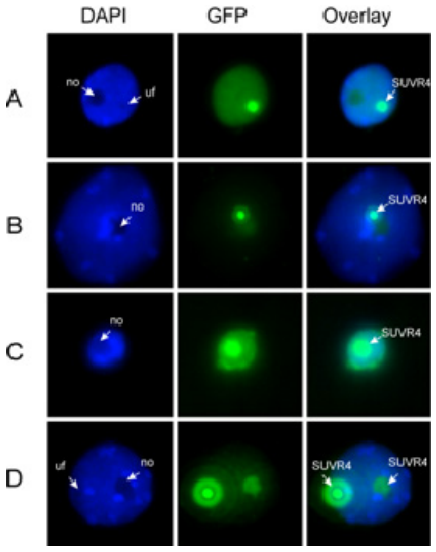

Figure S2

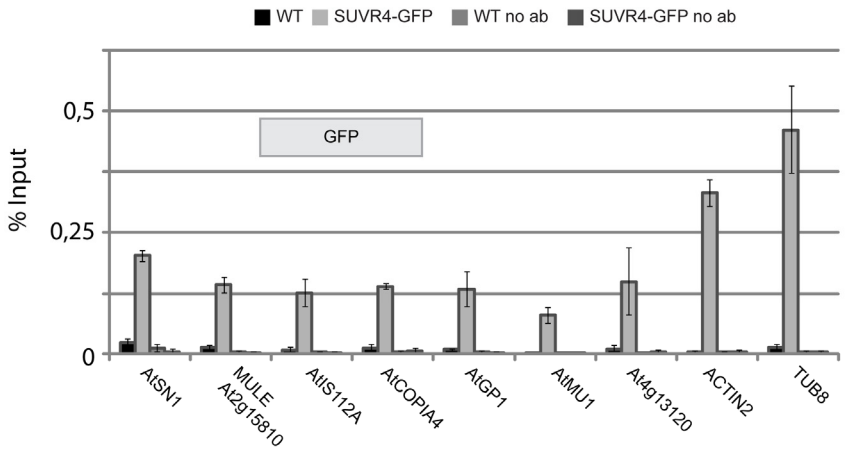
WT
SUVR4-GFP
WT no ab
SUVR4-GFP no ab
GFP
% Input
0,5
0,25
0
AtSN1
MULE At2g15810
AtIS112A
AtCOPIA4
AtGP1
AtMU1
At4g13120
ACTIN2
TUB8

Figure S3

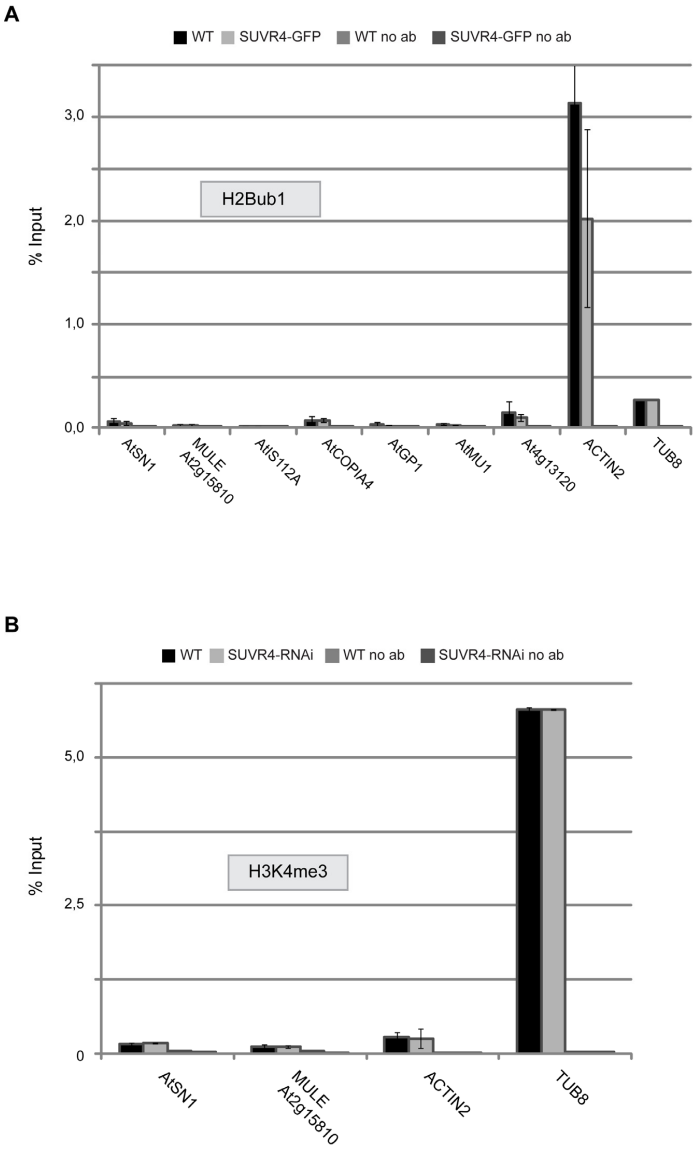
A
WT
SUVR4-GFP
WT no ab
SUVR4-GFP no ab
H2Bub1
% Input
3,0
2,0
1,0
0,0
AtSN1
MULE At2g15810
AtIS112A
AtCOPIA4
AtGP1
AtMU1
At4g13120
ACTIN2
TUB8
B
WT
SUVR4-RNAi
WT no ab
SUVR4-RNAi no ab
H3K4me3
% Input
5,0
2,5
0
AtSN1
MULE At2g15810
ACTIN2
TUB8

Figure S4

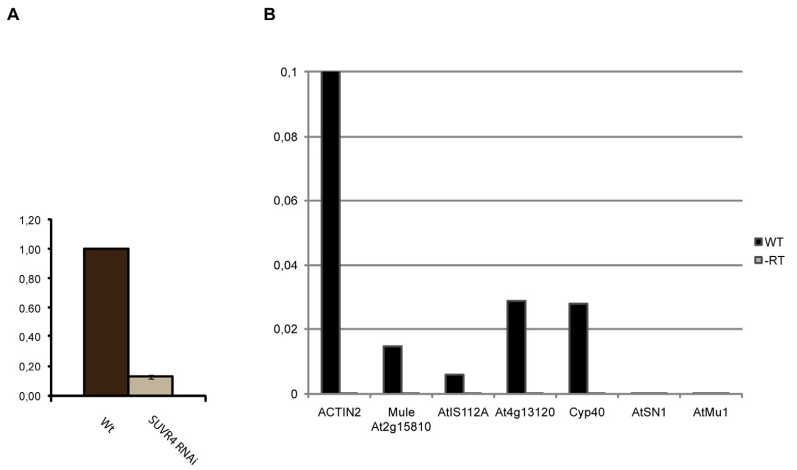

Figure S5

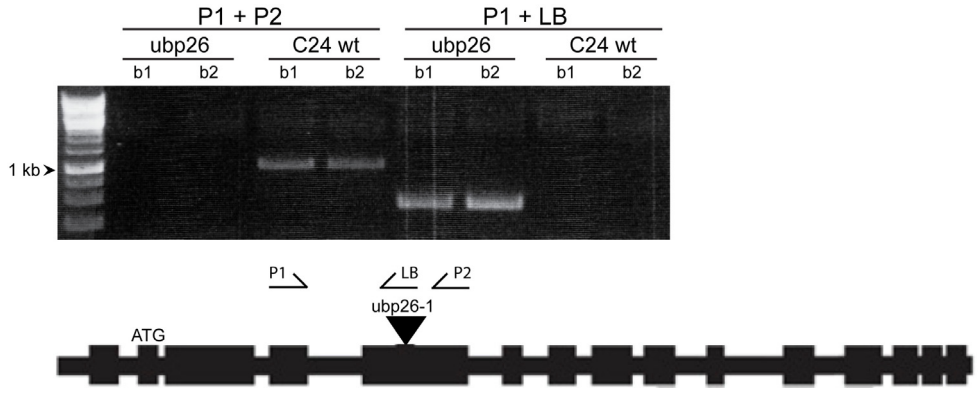

Figure S6

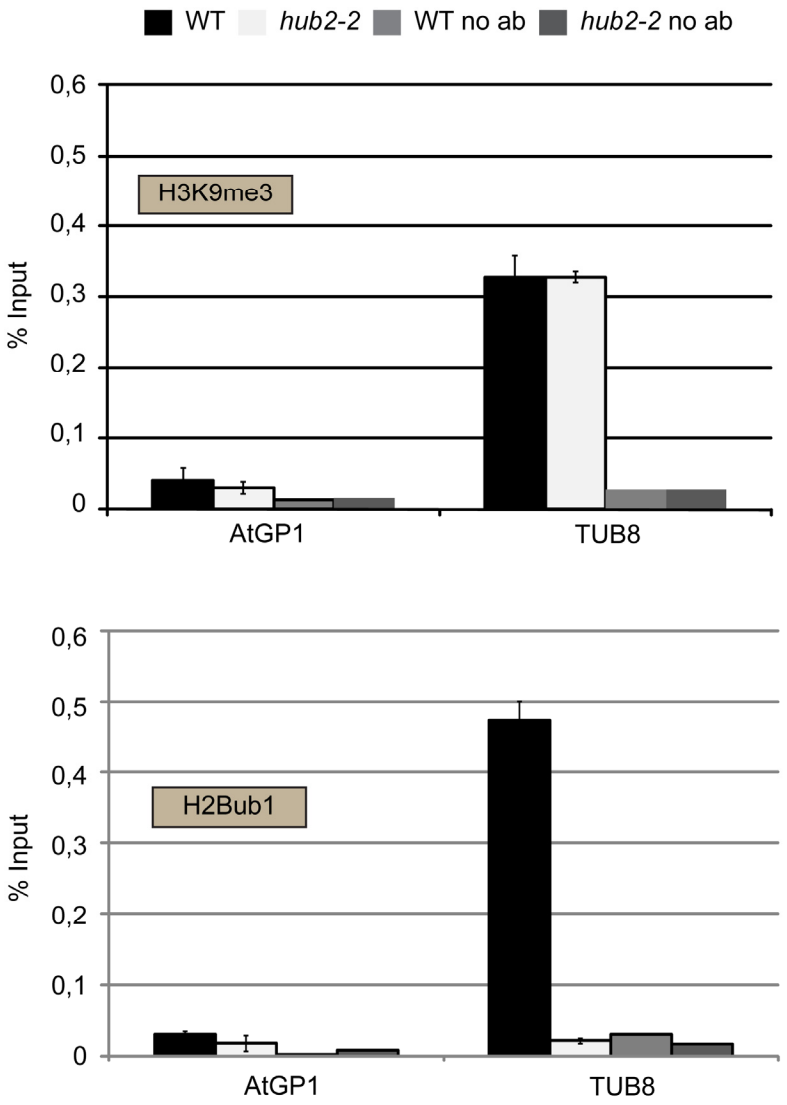
WT
hub2-2
WT no ab
hub2-2 no ab
H3K9me3
% Input
0,6
0,5
0,4
0,3
0,2
0,1
0
AtGP1
TUB8
H2Bub1
% Input
0,6
0,5
0,4
0,3
0,2
0,1
0
AtGP1
TUB8

Figure S7

Table S1. Oligos used in this study

Column1 Gene/clone	Column2 Name	Column3 Oligo Sequence	Column4 Application
TUBULIN8	TUB8 R	TGCAAATCGTTCTCTCCTTG	real time, ChIP
	TUB8 F	ATAACCGTTTCAAATTCTCTCTCTC	real time, ChIP
ACTIN2	act2int2_sense	CCCTGAGGAGCACCCAGTTCTACTC	real time, ChIP
	act2int2_antisense	CCGCAAGATCAAGACGAAGGATAGC	real time, ChIP
AtSN1	AtSN1-R	AAAATAAGTGGTGGTTGTACAAGC	real time, ChIP
	AtSN1-F	ACCAACGTGCTGTTGGCCCAGTGGTAAATC	real time, ChIP
	AtSN1me for	TTGGGGGGTTGTTAAAGTAGAGAA	Bisulphite
	AtSN1me rev	TTTTACTRCCATAAATCTTCTTC	Bisulphite
AtMU1	AtMU1-R	CTTAGCCTTCTTTTCAATCTCA	real time, ChIP
	AtMU1-F	GTGGATATACCAAAAACACAA	real time, ChIP
AtCOPIA4	AtCOPIA4-R	TGACGAAGAGCGTACCTGTG	real time, ChIP
	AtCOPIA4-F	CTTGTTTGTCTTCCCCGTGT	real time, ChIP
MULE At2g15810	MULE-R2	GATACTTGTTGACAAGTGTTTAGCAAGCC	ChIP
	MULE-F2	CTGTCCGCGAGTGTCATCAAGTAGC	ChIP
	2002-mule	TGTCTTCAGCTGCAGCATCATCAAC	real time
	2001-mule	TACAAGCTTCCAGAAGAGGAAATCTAT	real time
	At2g15810me rev	ATCATAACTTTTCAAARCTCTCAT	Bisulphite
	At2g15810me for	TTTGAAATGAGAATTGAAGGGTGAG	Bisulphite
Cyp40	Cyp40R	GATATATCCCCACCTTGTATC	real time
	Cyp40F	ATGGGTAGGTCAAAGTGTTTC	real time
AtIS112A	2g04293R	GCACAAGCTCTAAGATATTCTTGCTC	real time, ChIP
	2g04293F	CAAGAGGGATCATGTAGCACCAAAC	real time, ChIP
AtGP1	AtGP1-R	CAGAAAAATACTCGGTGCCAAT	real time, ChIP
	AtGP1-F	ACAGTGCCACAGTTGAGCAG	real time, ChIP
At4g13120	ASP/At4g13120	TTCTCTTGCAAATAGATCACAGC	real time, ChIP
	SP/At4g13120	ATACGACAATACTTTGTTCCAAAGG	real time, ChIP
SUVR4	SUVR4 854 L	CAGCTGACTTTCCTTGCACTTG	real time
	S4 1045 R	AACTTTCGGATTAAGTGTCC	real time
UBQ1	attB1 UBQ1 40-426	attB1-ATGCAGATCTTCGTGAAA	cloning
	attB2 UBQ1 40-426	attB2-CTACTTGATCTTCTTCTT	cloning
L40	attB1 UBQ1 268-426	attB1-ATTATTGAGCCTTCCTTGATGA	cloning
	attB2 UBQ1 40-426	attB2-CTACTTGATCTTCTTCTT	cloning
Ubiquitin	attB1 UBQ1 40-426	attB1-ATGCAGATCTTCGTGAAA	cloning
	attB2 UBQ1 40-267 STOP	attB2-CTAACCTCCTCTAAGCCTCAA	cloning
SUVR4-RNAi	SUVR4 RNAi BamHI 20 L	GACGGATCCACGACGCAGTGAAACAGAGA	cloning
	SUVR4 RNAi ClaI 405R	TATAACGACCAATTGCCACCA	cloning
	SUVR4 RNAi KpnI 405R	TATACGACCAATTGCCACCA	cloning
	SUVR4 RNAi XhoI 20L	ACGACGCAGTGAAACAGAGA	cloning
SUVR4-Full	S4-5UGWL	attB1-GTATGATCAGTCTCTCCGGACT	cloning
	S4GWR	attB2-TCATTTGCGCTTTTTAGACACCTC	cloning
SUVR4-WIYLD	S4-5UGWL	attB1-GTATGATCAGTCTCTCCGGACT	cloning
	S4-483 STOP GWR	attB2-TTATGAACCTTCACTTTGCTTATTC	cloning
R37	S4Wmut a109g_g110c_antisense	ATCGGGAATATCCAATTGCGCTGTTCTCTCTAAAGCTTTGAGTAC	mutagenesis
	S4Wmut a109g_g110c	GTACTCAAAGCTTTAGAGAGAACAGCGCAATTGGATATTCCCGAT	mutagenesis
D74	S4Wmut a221c	AACTATACTGCACTGGTCGCCGCTATTTATTCTGTTGAG	mutagenesis
	S4Wmut a221c_antisense	CTCAACAGAATAAATAGCGGCGACCAGTGCAGTATAGTT	mutagenesis
W61	S4Wmut a178t_t180a	CCTAGAAGAGGCTGGTGGCTAATGGTCGTATATAAAGTTGG	mutagenesis
	S4Wmut a178t_t180a_antisense	CCAACTTTATATACGACCATTAGCCACCAGCCTCTTCTAGG	mutagenesis
ubp26-1	F	GATTACCTTAGCTTGGAGCAGCTT	genotyping
	R	CTTCCTGATTTTCCTGTTTATCGCATCT	genotyping
T-DNA left boarder	LB	ACGCTGCGGACATCTACATTTTTGAATTG	genotyping

Cloning of DNA constructs: For yeast two-hybrid screening, the pDONR/Zeo-SUVR4-WIYLD construct was recombined into pGBKT7-GW [57], while pDONR/Zeo-UBQ1, pDONR/Zeo-ubiquitin and pDONR/Zeo-L40 were recombined into pGADT7 (Clontech) containing the Gateway recombination cassette. For protein expression, the following entry clones were recombined into pHMGWA [58] to make Maltose Binding Protein (MBP)-tagged proteins; pDONR/Zeo-SUVR4-Full, pDONR/Zeo-SUVR4-SACSET, pDONR/Zeo-UBQ1, pDONR/Zeo-Ubiquitin, pDONR/Zeo-L40. To create GST fusion proteins for pull-down, pDONR/Zeo-SUVR4-WIYLD was recombined into pGEX-AB-GW [28]. The pGEX-SUVR4-WIYLD clone was used as template for mutagenesis using the QuickChange II Site-Directed Mutagenesis Kit (Stratagene) according to the manufacturer's instructions with the following primers: S4W_R37A sense + antisense, S4W_W61R sense + antisense, S4W_D74A sense + antisense. pDONR/Zeo-SUVR4-Full was recombined into pEG104 [53] for overexpression of SUVR4-GFP in plants. The knock-down SUVR4 RNAi construct were made as follows: a fragment from the SUVR4 5'- end was amplified with two different primer pairs; PCR1 with SUVR4 RNAi BamHI 20L and SUVR4 RNAi ClaI 405R, and PCR2 with SUVR4 RNAi XhoI 20L and SUVR4 RNAi KpnI 405R. PCR product 1 was cut with restriction enzymes BamHI and ClaI, and ligated into the pHANNIBAL vector cut with the same enzymes. The resulting plasmid was amplified and purified, cut with XhoI and KpnI and ligated with XhoI- and KpnI-cut PCR product 2. The construct containing the SUVR4 fragments as inverted repeats with an intron in between, was subsequently cut out from pHANNIBAL using NotI, and finally ligated into the binary vector pART27 cut with NotI. All constructs were verified by sequencing.

II

III